

Eye-gaze Interaction Techniques for Use in Online Games and Environments for Users with Severe Physical Disabilities

Stephen Vickers
Department of Informatics
De Montfort University

A thesis submitted in partial fulfilment of the requirements of De Montfort
University for the degree of
Doctor of Philosophy (PhD)

June 2011

External Examiner:

Associate Professor John Paulin Hansen
Innovative Communication Group
IT University of Copenhagen, Denmark

Internal Examiner:

Professor Dimitrios Rigas
Faculty of Technology
De Montfort University, UK

Day of the defense: 9th May 2011

Abstract

Multi-User Virtual Environments (MUVES) and Massively Multi-player On-line Games (MMOGs) are a popular, immersive genre of computer game. For some disabled users, eye-gaze offers the only input modality with the potential for sufficiently high bandwidth to support the range of time-critical interaction tasks required to play. Although, there has been much research into gaze interaction techniques for computer interaction over the past twenty years, much of this has focused on 2D desktop application control. There has been some work that investigates the use of gaze interaction as an additional input device for gaming but very little on using gaze on its own. Further, configuration of these techniques usually requires expert knowledge often beyond the capabilities of a parent, carer or support worker.

The work presented in this thesis addresses these issues by the investigation of novel gaze-only interaction techniques. These are to enable at least a beginner level of game play to take place together with a means of adapting the techniques to suit an individual. To achieve this, a collection of novel gaze based interaction techniques have been evaluated through empirical studies. These have been encompassed within an extensible software architecture that has been made available for free download. Further, a metric of reliability is developed that when used as a measure within a specially designed diagnostic test, allows the interaction technique to be adapted to suit an individual. Methods of selecting interaction techniques based upon game task are also explored and a novel methodology based on expert task analysis is developed to aid selection.

Acknowledgements

I would like to acknowledge all of my co-authors and collaborators that have made this work possible. Also, a special thank you to the Institute of Creative Technologies and Andrew Hugill that have have funded my PhD.

A thank you to my three supervisors Howell, Aulikki and Bob; all of which have offered me valuable support and assistance over the last few years. Howell and Aulikki have steered me along the path of gaze interaction and gaming and have shown me how research should be done. They also made it possible for a collaboration with the Visual Interaction Research Group (VIRG) at the University of Tampere where important parts of this work was conducted. Also, a further special thank you to Howell who has always been available if I needed help; and motivational if I needed encouraging. Further, a thanks to Richard, Santtu and Lauri who in addition to being co-authors for papers have helped with software development.

A thank you to Jane, Helen, Barry and all of the children at Ash Field school who despite being very busy have always made time for me to visit.

I would also like to thank all of the staff and PhD students within the Centre for Computational Intelligence for their help in debugging software (especially L^AT_EX!), discussions and for taking part in pilot studies.

I have had the opportunity to participate in conferences, visit a number of institutions and meet lots of great people. This is due to much appreciated funding from COGAIN and De Montfort University.

A further thank you to my two examiners John and Dimitrios for reviewing my thesis and providing an enjoyable examination.

Finally, it would not have been possible for me to make the commitment that a PhD requires without the patience and support from my wife Abi.

Declaration

The work described in this thesis is the original work of the author except where specific reference or acknowledgement is made to the work or contribution of others. The following conference papers and journal articles have contributed to the research within the thesis with some portions reproduced as they are and others revised and edited. All authors have made valuable contributions to these publications and I have referred to my specific role at the beginning of each chapter where they are used.

Bates, R.; Vickers, S. & Istance, H. **Gaze Interaction with Virtual On-Line Communities: Levelling the Playing Field for Disabled Users.** *Universal Access in the Information Society; UAIS*. SpringerLink, 2009. 9. 261-272

Vickers, S.; Bates, R. & Istance, H. **Gazing into a Second Life: Gaze-Driven Adventures, Control Barriers, and the Need for Disability Privacy in an Online Virtual World.** *Proceedings of the 7th International Conference on Disability, Virtual Reality and Associated Technologies; ICDVRAT 2008*. 2008

Istance, H.; Bates, R.; Hyrskykari, A. & Vickers, S. **Snap Clutch, a Moded Approach to Solving the Midas Touch Problem.** *Proceedings of the 2008 Symposium on Eye Tracking Research and Applications; ETRA 2008*, 2008

Vickers, S.; Istance, H.; Hyrskykari, A.; Ali, N. & Bates, R. **Keeping an Eye on the Game: Eye Gaze Interaction with Massively Multiplayer Online Games and Virtual Communities for Motor Impaired Users.** *Proceedings of the 7th International Conference on Disability, Virtual Reality and Associated Technologies; ICDVRAT 2008.* 2008

Istance, H.; Hyrskykari, A.; Vickers, S. & Chaves, T. **For Your Eyes Only: Controlling 3D Online Games by Eye-Gaze.** *Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: INTERACT 2009.* 2009

Vickers, S.; Istance, H. & Hyrskykari, A. **Selecting Commands in 3D Game Environments by Gaze Gestures.** *5th Conference on Communication by Gaze Interaction; COGAIN.* 2009

Istance, H.; Vickers, S. & Hyrskykari, A. **Gaze-based Interaction with Massively Multiplayer On-line Games.** *CHI EA '09: Proceedings of the 27th International Conference Extended Abstracts on Human Factors in Computing Systems.* 2009

Vickers, S.; Istance, H. & Smalley, M. **EyeGuitar: Making Rhythm Based Music Video Games Accessible Using Only Eye Movements.** *Proceedings of the 2010 International Conference on Advances in Computer Entertainment Technology: ACE '10.* 2010

Contents

List of Figures	ix
List of Tables	xiv
I Introduction and Background	1
1 Introduction	2
1.1 Motivation	2
1.2 Research Directions	6
1.3 Thesis Structure	7
2 The Eye, Eye Tracking and Physical Disabilities	10
2.1 The Eye	10
2.1.1 Human Vision	11
2.1.2 Characteristics of Eye Movements	14
2.2 Eye and Gaze Tracking	15
2.2.1 Non-invasive Eye Tracking	16
2.2.2 Invasive Eye Tracking	17
2.2.3 Open Source Eye Tracking	18
2.3 Physical Disabilities	19
2.3.1 Augmentative and Alternative Communication (AAC)	20
2.4 Implications for Eye and Gaze Tracking	22
2.5 Summary	26

3	Gaze Based Interaction with Desktop Environments	27
3.1	Gaze Interaction Techniques	27
3.1.1	Fundamentals of Gaze Interaction	28
3.1.2	Application Interface Control	29
3.1.3	Text Entry and Communication	33
3.1.4	Drawing	38
3.1.5	Reading and Internet Browsing	41
3.1.6	Gaming	42
3.2	Adaptive Interfaces	45
3.3	Summary	48
II	Gaze Interaction Techniques for Game Tasks	49
4	Initial Assessment of Gaze Interaction with Online Games	50
4.1	Multi-User Virtual Environments (MUVE) and Massively Multiplayer Online Games (MMOG)	51
4.2	Task Interaction	54
4.3	Pilot Study 1 - Initial Assessment of Gaze Interaction	55
4.3.1	Participants and Apparatus	56
4.3.2	Procedure	56
4.3.3	Analysis	57
4.4	Discussion and Conclusions	60
5	Bottom Up Approach: Gaze Interaction Techniques	62
5.1	A Software Device for Interpreting Gaze Data	63
5.1.1	Modes of Interaction	65
5.1.2	Architecture and Implementation	65
5.2	Task Categories and Novel Gaze Interaction Techniques	68
5.2.1	Locomotion	69
5.2.2	Camera Movement	71
5.2.3	Object Manipulation and Application Control	72
5.2.4	Communication	74
5.3	Modifying Gaze Interaction Techniques	76

5.4	Evaluation Study 1: Gaze Interaction in Second Life	80
5.4.1	Participants and Apparatus	81
5.4.2	Procedure	81
5.4.3	Analysis and Results	83
5.4.4	Conclusions	86
5.5	Evaluation Study 2: Gaze Interaction in World of Warcraft	86
5.5.1	Participants and Apparatus	87
5.5.2	Procedure	87
5.5.3	Analysis and Results	88
5.5.4	Evaluation Study 1 and 2 Discussion	91
5.6	Discussion and Conclusions	93
6	Game Task Analysis	95
6.1	Typical Tasks in an MMORPG	96
6.2	A Methodology for Selecting Interaction Techniques based on Game Tasks	98
6.2.1	Task Analysis	98
6.2.2	Evaluation Study 3A: Task Analysis of an MMORPG	99
6.2.3	Task Modelling of the Locomotion Tasks	104
6.2.4	Object-Oriented Task Domains	113
6.2.5	Selecting an Appropriate Gaze Interaction Technique	120
6.2.6	Evaluation Study 3B: Comparing Event Signatures Between Gaze and Keyboard	123
6.3	EyeGuitar	124
6.4	Concluding Summary	127
III	User Evaluation	128
7	Dynamic Accessibility	129
7.1	Dynamic Accessibility	129
7.2	User Groups	130
7.3	Experiment 1: Understanding User Requirements	134
7.3.1	Configuration	134
7.3.2	Procedure	136

7.3.3	Observations	137
7.4	Determining Performance of a Gaze Interaction Technique	140
7.4.1	Influencing Factors	142
7.4.2	A Diagnostic Test Based Upon the Interaction Technique Re- quirements	144
7.4.3	Developing a Metric of Reliability (R_{user})	146
7.4.4	Adapting the Interaction Technique	149
7.4.5	Non-conventional Design	154
7.4.6	Fitts' Law	154
7.5	Experiment 2: Diagnostic Test and Adaption of the Interaction Technique	155
7.5.1	Participants and Configuration	155
7.5.2	Procedure	155
7.5.3	Analysis and Results	155
7.6	Experiment 3: Validating the Adapted Interaction Technique	162
7.6.1	Configuration	163
7.6.2	Procedure	163
7.6.3	Orienteering Task	164
7.6.4	Analysis and Results	165
7.7	Discussion and Conclusions	174
8	Evaluation and Comparison Between all Group Participants	177
8.1	Experiment 4: Determining User Reliability (R_{user})	177
8.1.1	Participants and Configuration	177
8.1.2	Procedure	178
8.1.3	Results and Analysis	178
8.2	Experiment 5: Locomotion Task	182
8.2.1	Participants and Configuration	184
8.2.2	Procedure	185
8.2.3	Results and Analysis	185
8.3	Discussion and Conclusions	190
9	Conclusions	193
9.1	Contributions	193
9.2	Critical Review and Future Work	197

IV	Appendices	200
A	Experiment 2, Supporting Data	201
B	Experiment 4, Supporting Data	206
	B.1 Notes on Resampling	206
C	Experiment 5, Supporting Data	210
	References	215

List of Figures

1.1	Disability privacy burden	4
1.2	The approach used in attempting to generate a complete gaze interface.	6
1.3	Structure of the thesis	9
2.1	Diagram of the eye	11
2.2	Density of rod and cone receptors	12
2.3	Visual angle	12
2.4	Visual angle subtended by the fovea	13
2.5	Purkinje reflections	17
2.6	Pupil and first Purkinje reflection positions	17
2.7	Electro-oculography (EOG)	18
2.8	Etran frame	21
2.9	AAC dynamic screen communication device	22
2.10	AAC switches	23
2.11	A young cerebral palsy sufferer sitting in his natural divergent position.	25
3.1	Gesture recogniser directions	32
3.2	Example gaze gestures	32
3.3	On screen keyboard	34
3.4	GazeTalk's full screen keyboard	34
3.5	The Quikwriting interface adapted for use with eye gaze	35
3.6	EyeWrite's letter chart	36
3.7	EyeWrite being used with Notepad	37
3.8	Using StarGazer to select the letter 'S'	38
3.9	Dasher	39

LIST OF FIGURES

3.10	EaglePaint	40
3.11	EyeDraw	40
3.12	Automatic text scrolling	42
3.13	Single strok gaze gestures	44
3.14	Task types used in SUPPLE++	46
3.15	Generated user interfaces based on SUPPLE	47
4.1	Screenshot of Second Life	52
4.2	Screenshot of World of Warcraft	52
4.3	Screenshot of Free Realms	53
4.4	Second Life Control Panels	55
4.5	Pilot Study 1: Locomotion and distraction	58
4.6	Pilot Study 1: Camera movement and feedback	59
4.7	Pilot Study 1: Object manipulation	60
4.8	Pilot Study 1: Feedback while eye typing	61
5.1	Snap Clutch: Interface	66
5.2	Snap Clutch: Operation	67
5.3	‘Drag-from-here’ cursor.	70
5.4	‘Look around’ task mode	70
5.5	‘Locomotion’ task mode.	71
5.6	‘Park-it-here’ cursor.	72
5.7	Snap Clutch: Magnifying toolglass	73
5.8	Snap Clutch: Magnifying glass	73
5.9	Performing a gesture in World of Warcraft	74
5.10	Snap Clutch: Gesture configuration	75
5.11	Snap Clutch: G9 predictive text	76
5.12	Variants based on complexity choices	79
5.13	Variants based on complexity functions	79
5.14	Variants based on accuracy requirements	79
5.15	Variants based on temporal requirements	79
5.16	Transparent overlay variations	80
5.17	Evaluation Study 1: Locomotion task	82
5.18	Evaluation Study 1: Results	85

LIST OF FIGURES

5.19	Evaluation Study 2: Results	89
6.2	Evaluation Study 3A: Starting Position	100
6.3	Evaluation Study 3A: Participant heat maps 1	102
6.4	Evaluation Study 3A: Participant heat maps 2	102
6.5	Evaluation Study 3A: Participant heat maps 3	102
6.6	Evaluation Study 3A: Time-line of tasks 1	103
6.7	Evaluation Study 3A: Time-line of tasks 1	103
6.8	Event signature example	104
6.9	GOMS: Locomotion task using keyboard	106
6.10	GOMS: Locomotion task using gaze	106
6.11	BNF: Locomotion task using keyboard	107
6.12	BNF: Locomotion task using gaze	107
6.13	UAN: Example	108
6.14	UAN: Locomotion task using keyboard	109
6.15	UAN: Locomotion task using, gaze	109
6.16	ExperiScope symbols	110
6.17	ExperiScope example	111
6.18	ExperiScope: Locomotion task using keyboard	112
6.19	ExperiScope: Locomotion task using gaze	112
6.20	Three task classifications	113
6.21	AHP Hierarchy	119
6.22	Transparent overlay with heat map	122
7.1	A scale of eye-gaze as a requirement for primary communication.	130
7.2	Group A participant walking around in World of Warcraft	132
7.3	Plan view of the experiment setup used	135
7.4	Example of a poor calibration	137
7.5	Screenshot of the diagnostic test	145
7.6	Diagnostic test in progress.	145
7.7	Re-factoring time as a percentage to calculate R_{time}	147
7.8	Re-factoring time as a percentage to calculate $R_{fixation}$	148
7.9	The priority levels of character control	150

LIST OF FIGURES

7.10	Zone and function positions for the transparent overlay interaction technique	151
7.11	Priority system process flow	153
7.12	Example of an adapted technique	154
7.13	Experiment 2: Comparing ‘success’ as a measure with reliability	157
7.14	Example surface map	158
7.15	Experiment 2: Surface map and adapted interaction technique 1	158
7.16	Experiment 2: Surface map and adapted interaction technique 2	159
7.17	Experiment 2: Surface map and adapted interaction technique 3	159
7.18	Experiment 2: Surface map	160
7.19	Experiment 2: Successful gaze scan path	160
7.20	Experiment 2: Gaze scan paths 1	161
7.21	Experiment 2: Gaze scan paths 2	161
7.22	Experiment 2: Gaze scan paths 3	162
7.23	Modified interaction technique	164
7.24	Experiment 3: Map overview	165
7.25	Experiment 3: Start position	166
7.26	Experiment 3: Waypoint 1	166
7.27	Experiment 3: Waypoint 2	167
7.28	Experiment 3: Waypoint 3	167
7.29	Experiment 3: Participant performing the orienteering task	168
7.30	Experiment 3: Participant performing the orienteering task 2	168
7.31	Experiment 3: Participant performing the orienteering task 3	169
7.32	Experiment 3: Route taken by GP-A2	170
7.33	Experiment 3: Route taken by GP-A1	171
7.34	Experiment 3: Route taken by GP-A3	171
7.35	Experiment 3: Route taken by GP-A4	172
7.36	Experiment 3: Results	173
8.1	Experiment 4: Reliability by user group	179
8.2	Experiment 4: Reliability (R_{user}) distribution of groups B and Group C	180
8.3	Experiment 4: Fixations made vs time to complete	181
8.4	Experiment 4: Time eyes not visible vs time to complete	181

LIST OF FIGURES

8.5	Distribution of fixations made	182
8.6	Distribution of completion time	183
8.7	Distribution of time eyes are not being tracked	183
8.8	Experiment 5: Surface map and adapted interaction technique 1	184
8.9	Experiment 5: Surface map and adapted interaction technique 2	185
8.10	Experiment 5: Time to reach the first waypoint by user group	187
8.11	Experiment 5: Distribution of time in reaching the first waypoint	188
8.12	Experiment 5: Total time to reach all waypoints for MD participants in Group B and AB participants in Group C	189
8.13	Time to reach the first waypoint versus reliability	192
9.1	Gesture based diagnostic test	199
A.1	Results from Experiment 2: GP-A1	202
A.2	Results from Experiment 2: GP-A2	203
A.3	Results from Experiment 2: GP-A3	204
A.4	Results from Experiment 2: GP-A4	205
B.1	Experiment 4: Data Resampled 10,000 times from Groups A and B . . .	207
B.2	Experiment 4: Data Resampled 10,000 times from Groups A and C . . .	207
B.3	Experiment 4: Data Resampled 10,000 times from Groups B and C . . .	208
B.4	Reliability values from the diagnostic tests in Experiment 2 and Exper- iment 4 for all participant Groups.	209
C.1	Experiment 5: Data Resampled 10,000 times from Groups A and B . . .	210
C.2	Experiment 5: Data Resampled 10,000 times from Groups A and C . . .	212
C.3	Experiment 5: Data Resampled 10,000 times from Groups B and C . . .	212
C.4	Time to reach the first waypoint from Experiment 3 and Experiment 5 for all participant Groups.	213
C.5	Task completion times for all Groups with reliability overlaid	214
C.6	Experiment 5: Data Resampled 10,000 times from Groups B and C. This is based on total time.	214

List of Tables

3.1	Taxonomy of gaze interaction technique applications	30
3.2	Performance comparison of different gaze based gesture implementations	45
4.1	Control requirements for task categories	55
4.2	Pilot Study 1: Results	57
5.1	Example of interaction technique variants	78
5.2	Evaluation Study 1: Results	84
5.3	Pilot Study 3: Results	89
6.1	Task breakdown of open play in an MMORPG over two sessions	97
6.2	Evaluation Study 3A: A sample summary of the transcript taken from a participant during the experiment	101
6.3	Interaction notation structure	114
6.4	Basic set of interaction notation	115
6.5	Interaction notation: Locomotion task using keyboard	115
6.6	Interaction notation: Locomotion task using Snap Clutch	116
6.7	Interaction notation: Actual task model	117
6.8	Task requirements and technique properties	118
6.9	Evaluation Study 3A: Summary of the game tasks and candidate gaze interaction techniques	121
7.1	Summary of the participants in Group A	133
7.2	Experiment 1: Observation summary	141
7.3	Definition of attribute levels	142
7.4	Influences on gaze interaction performance	142

LIST OF TABLES

7.5	Overview of the priority system to which functions are assigned to zones	151
7.6	Example of the priority selection system	152
7.7	Experiment 2: Results	156
7.8	Experiment 3: Results 1	172
7.9	Experiment 3: Results 2	173
8.1	Reliability (R_{user}) results for all three participant groups	178
8.2	Experiment 5: Time for all three participant groups. Group A contains CP participants; Group B contains MD participants; Group C contains AB participants.	186
8.3	Experiment 5: Total time for participant Groups B and C. Group B contains MD participants; Group C contains AB participants.	188
8.4	The greater the reliability then the better task time performance when sorting by group median.	191
C.1	Experiment 3 and 5: Task completion times	211

Part I

Introduction and Background

Chapter 1

Introduction

1.1 Motivation

Video games have grown to become one of the most popular forms of entertainment alongside watching movies and listening to music. One reason for this is that they are one of the few forms of entertainment that are truly interactive. They are designed and programmed to evoke a sense of accomplishment and satisfaction when they are played. There is also no limit or restriction on what constitutes a video game and this has led to a countless list of genres being created. This flexibility with genre and the type of gameplay involved has partly contributed to gaming becoming a more accepted form of entertainment. A gamer can race in the fastest car, play in a virtual band or build an entire city. In contrast, they can also learn how to design and engineer their car to become faster or more fuel efficient, learn about rhythm and music or learn about the importance of economics and the influence of politics in city management. Gaming has a much larger role in society than just entertainment and is becoming increasingly important in education and social media. In recent years, online gaming has expanded to include almost all genres of gaming. All of which share the common element of them not only being online but also requiring interaction with other players.

The work in this thesis is applied to the genre of Multi-User Virtual Environments (MUVEs) and Massively Multi-player Online Role Playing Games (MMORPG). These immersive, realistic video games have become incredibly popular. The MUVE of Second Life, for example, has over 14 million members (Linden-Labs, 2011) and can support interest groups, cooperative activities between people as well as serious commercial

activities. In contrast, the MMORPG World of Warcraft is more focused on character development achieved through completing game related goals. World of Warcraft has more than 12 million users worldwide (Blizzard, 2010). There are also MMORPG's designed specifically for children in which, character development also involves completing educational elements such as maths, reading and language skills (e.g. Free Realms and Chobots).

People with severe physical disabilities can derive much enjoyment from playing these games. Participation can be challenging, fun and educational, as well as giving opportunities for social interaction. However, if their disability means that they are unable to use a conventional mouse, keyboard or game-pad then they will now be presented with a barrier of how they might interact with the game. Different games require different levels of interaction to be performed at different speeds, some of them in real-time. Although, many augmentative and alternative communication devices exist as replacements to mouse, keyboard and game-pads, they are often too slow and cumbersome to use.

Control Demands

As online games strive for more realism, the levels of interaction required increases making accessibility more critical for disabled users. Figure 1.1, highlights the increased difficulty as the demands of communication time (the latency of your response to a communication or action from another online person) and the amount of data that needs to be generated to enable complete communication or control over your online presence increases.

It is this ever increasing control bandwidth that is giving rise to avatar realism that contributes to the increasing need for communication and control efficiency. The move to realism and feature-rich interaction can be seen by considering four different 'chat' methods; *email*, *chat rooms*, *simple avatars* and *realistic avatars*. At the very basic level of email there is no time pressure that the user is faced with - they can take several hours, days or even weeks to respond. A chat room requires faster interaction from the user than is required for email but it is acceptable to have tens of seconds or maybe minutes to respond. An MUVE like Second Life comes under the simple avatar category and requires real-time interaction in order to participate fully. Many users expect almost instant responses when chatting at this level and not to meet this demand

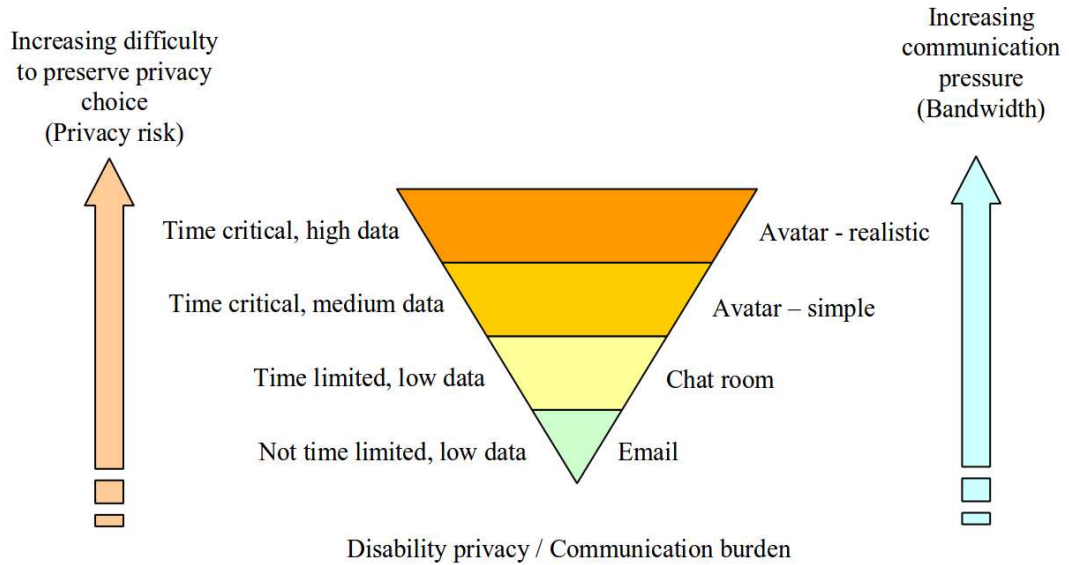


Figure 1.1: Disability privacy burden by online meeting type (Bates et al., 2008)

can result in users becoming bored and questioning why the avatar is not responding like everyone else.

Many online games have already begun to move towards creating realistic avatars. In this category, avatars are able to verbally communicate using VoIP (Voice Over IP) technology. There are also possibilities for the avatar to respond to our facial expressions and body movements due to advancements in facial and gesture tracking. This hierarchy of an ever increasing communication burden on disabled users threatens their ability to fully participate.

Eye-gaze Interaction

For some groups of people, eye-gaze offers the only input modality with the potential for sufficiently high bandwidth to support the range of time-critical interaction tasks required to participate in these games. Gaze based interaction began in the early 1980's (e.g. Friedman et al., 1982; Kate et al., 1979) and since then has been shown as an effective means of computer control for users with a severe physical disability (e.g. Bates, 2002; Hansen et al., 2006). There has been much research conducted into 2D desktop interaction (e.g. Lankford, 2000; Skovsgaard et al., 2010; Zhai et al., 1999) and "eye-typing" (text entry using only eye movements) (e.g. Johansen et al.,

2003; Majaranta and Raiha, 2002; Ward and Mikaelian, 2002) and much of it has been successful in advancing the field. There has been some research into using eye-gaze as an additional input device for gaming (e.g. Isokoski and Martin, 2006; Smith and Graham, 2006) but little on using gaze as the only input modality. Much of this work has considered able-bodied gamers and how gaze could be integrated with mouse, keyboard and game-pad. Using gaze as a single input device for game control for people with disabilities who are only able to use eye-gaze, presents a different set of challenges of which have only just begun to be addressed. Many of the existing gaze interaction techniques rely on the user performing unnatural eye movements in order to initiate command sequences. This may include looking at an object on screen for a prolonged period of time or perhaps making a patterned set of eye movements. However, this unnatural, explicit, command-based method of control can be costly in terms of time, effort and errors made. Gaze interaction for video games requires different interaction techniques and more consideration of the range and speed of the tasks being performed.

Eye tracking technology can cost several thousands of pounds and so is beyond the reach of many who can benefit from it. However, the greater the interest that is generated, then the greater the reduction in cost due to economies of scale (Hansen et al., 2005) and also the increased availability of open source alternatives (San Agustin et al., 2009). As eye tracking does become more available the demand for more efficient and configurable gaze interaction techniques will also increase. There is often a phenomenalist approach to gaze interaction configuration for disabled users (Donegan et al., 2005), in that each person is considered unique and no generalisations can be made about which interaction techniques best suit different types of disabilities. This results in many requiring expert configuration of gaze controlled user interfaces so as to suit an individual. The expert knowledge required is likely to be beyond the capabilities of a parent, carer or support worker (Dawe, 2006). So, in order for gaze interaction to be widely used there is a need for it to become less reliant upon resource intensive, expert configuration.

This thesis aims to provide the basis for automatic or semi-automatic generation of gaze driven interfaces for individuals wishing to play different games. The approach used requires knowledge about the interaction techniques being performed, the requirements of the game tasks, the capabilities of the eye tracker and the requirements and limitations of the user, see Figure 1.2. All of these components have an influence on

the selection of an appropriate interaction technique for a given game task. To cover all genres of game in this thesis is not possible and so the focus of this research is on MUEs and MMORPGs. However, the theory and ideas are intended to be a platform for extending the approach to all genres of game.

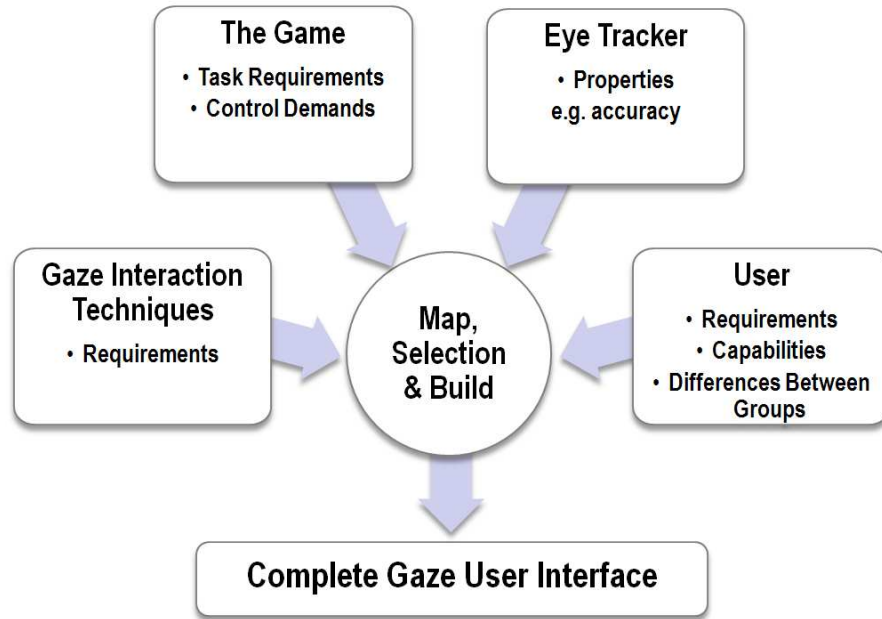


Figure 1.2: The approach used in attempting to generate a complete gaze interface.

1.2 Research Directions

The main research question addressed by this thesis is:

Can an eye-gaze only interface be automatically generated to suit the specific needs of a person with severe physical disabilities so that they may fully participate in an immersive computer game?

To answer this question the following objectives are to be met. Firstly, existing and new gaze interaction techniques need to be evaluated:

Can the typical range of game tasks be completed using gaze-only interaction techniques?

How can these techniques be adapted if required?

Secondly, there is a need to develop a mechanism for selecting candidate eye-gaze interaction techniques based upon a set of game tasks and the needs of the user:

What are the task characteristics and how does this affect the requirements of the interaction technique?

Finally, there is a need to evaluate the needs of individual users:

How well can a particular gaze interaction technique be used by an individual with their own capabilities and limitations?

Can we begin to make generalisations about differences between disability groups and how these vary from able-bodied groups?

1.3 Thesis Structure

This thesis is structured into three main parts, see Figure 1.3.

Part I: Introduction and Background

This part presents the motivation for the research and defines the research questions. It describes the physiology of the eye and the technology used to determine its movement and position. This leads into an overview of physical disabilities and the role of augmented and alternative communication. The part concludes with a review on the different methods used to enable gaze interaction with 2D desktop and gaming environments.

Part II: Gaze Interaction Techniques for Game Tasks

The second part of the work involves a pilot study and several empirical evaluation studies with able-bodied participants. These were all controlled studies that allowed interaction techniques, software and experimental tasks to be properly defined prior to further trials with participants with disabilities. *Chapter 4* introduces the virtual environments and games used throughout the thesis and presents a pilot study on gaze based interaction within an MUVE. *Chapter 5* extends the work in the previous chapter by the development of new interaction techniques and their evaluation along with existing techniques. This allows for a taxonomy of gaze interaction technique

modification properties to be developed. *Chapter 6* describes a methodology for a detailed task analysis and model based on expert gameplay. This helps to define the task properties and the subsequent requirements that the interaction technique is required to fulfil. An evaluation study is conducted using an MMOG and the results allow several candidate gaze interaction techniques to be matched to specific game tasks.

Part III: User Evaluation

The final part of the thesis is a series of experiments with three different participant groups. *Chapters 7 and 8* begin with an observational experiment that was intended to see what the range of tasks disabled participants could complete. However, participants found that the interaction techniques were difficult to use even though able-bodied participants had previously been able to use them. Therefore, the direction of research evolved into being an investigation into why this was so and subsequently a metric of reliability and a diagnostic test based upon one of the interaction techniques was developed. The results of the diagnostic tests allow for a visualisation model of reliability to be created and suggest how the interaction technique could be adapted to better suit a users ability. Further experiments then verified that the adapted interaction technique improved the performance of the participants when performing a locomotion based task. *Chapter 8* attempts to describe generalisations about differences between different disability groups based upon the reliability metric and task performance. Finally, *Chapter 9* summarises and concludes the findings and contributions made.

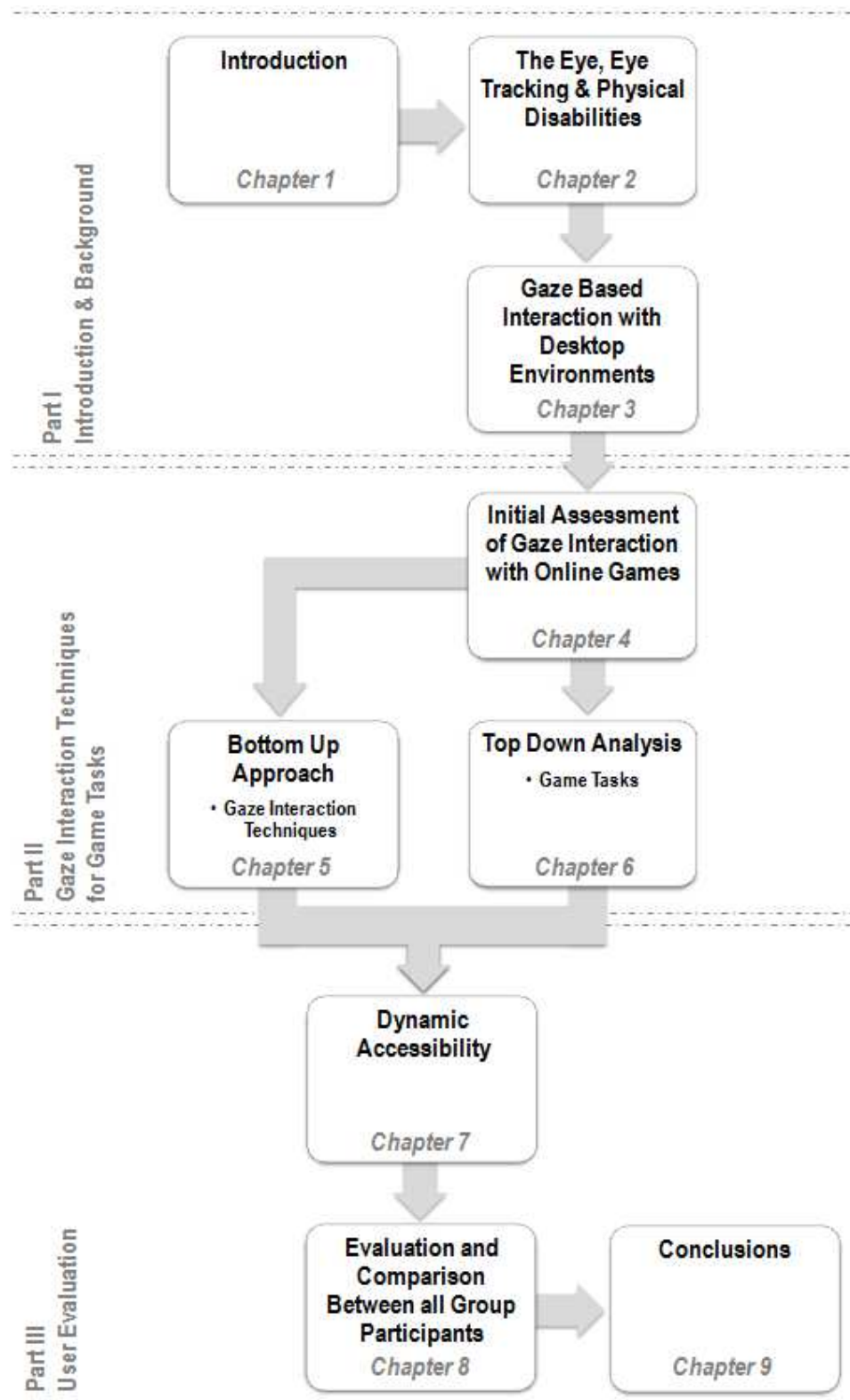


Figure 1.3: Structure of the thesis

Chapter 2

The Eye, Eye Tracking and Physical Disabilities

The aim of this Chapter is to give an overview of eye anatomy and movements, the technology behind eye tracking and physical disabilities. It begins with a brief explanation on human vision and the different types of movements that are made. This leads on to the alternative methods used for eye tracking, including commercial and open source eye tracking systems. The Chapter then goes on to give a brief overview of physical disabilities and how Augmentative and Alternative Communication (AAC) is used to support or replace verbal communication. Finally, the Chapter concludes with a discussion on the implications of measuring eye movements in regards to eye anatomy and movements, technology and body movement.

2.1 The Eye

The studying of eye movements goes back to the late 19th century when scientists such as Louis Émile Javal (Javal, 1907) first tried to describe the way in which people read. Javal relied on his own eye observations of readers to determine that eyes do not move in a straight line when reading text but rather make short rapid movements (saccades) with short pauses (fixations) in between. Since these early observations scientists have been interested with the way that we move our eyes in order to read, observe the world and look at each other.

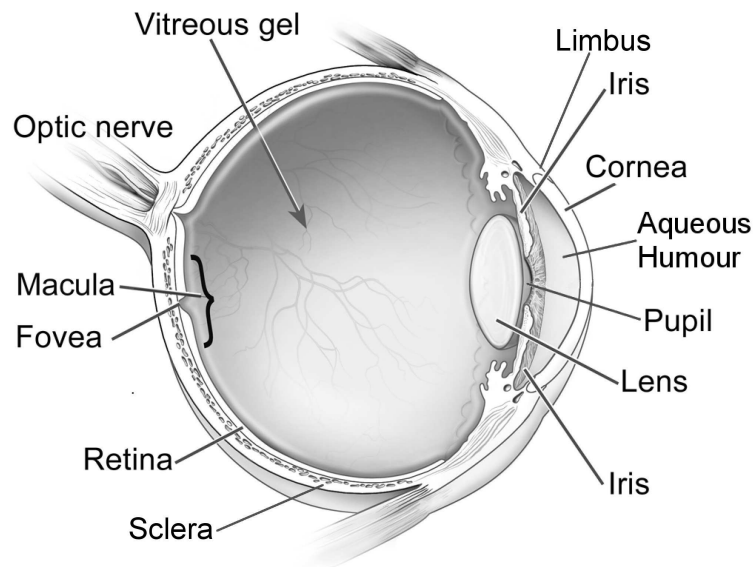


Figure 2.1: Diagram of the eye (*Eye diagram*, 2010)

2.1.1 Human Vision

The eye is a complex organ that we use to view the world, see Figure 2.1. The outermost layer, called the sclera holds the eyes shape and at the front most part of this is the cornea. The area between the sclera and cornea is known as the limbus. The transparent cornea is the window into the eye that refracts light as it passes through before reaching the aqueous humor, lens, vitreous humor and finally the retina. The amount of light that enters is controlled by the pupil which expands and contracts accordingly. The retina contains two types of photoreceptor cells: rods and cones. The rods are highly sensitive to light and used in conditions where light is less intense and as such are unable to cope with fine detail often becoming saturated with light; as can be experienced when moving from a dark room to a brightly lit room. There are approximately 120 million rods per eye distributed across the retina as seen in Figure 2.2. The cone photoreceptors are much smaller in number with approximately 6 million per eye. These are mainly concentrated on a small part of the retina known as the fovea which is responsible for sharp vision. The Figure shows the contrast in quantity and visual angle between the two types of photoreceptor cells.

The visual angle α (see Figure 2.3) is the angle of which light reflected from an object is focused through the lens onto the surface of the retina, where D is distance

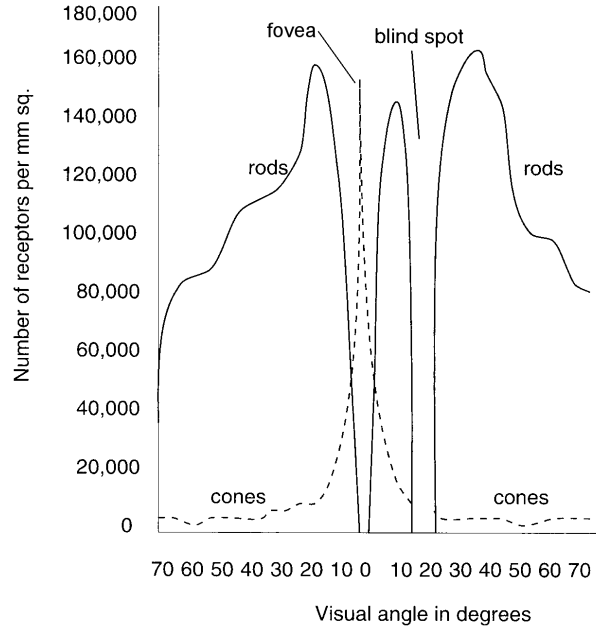


Figure 2.2: Density distributions of rod and cone receptors across the retinal surface. Adapted from Duchowski (2003).

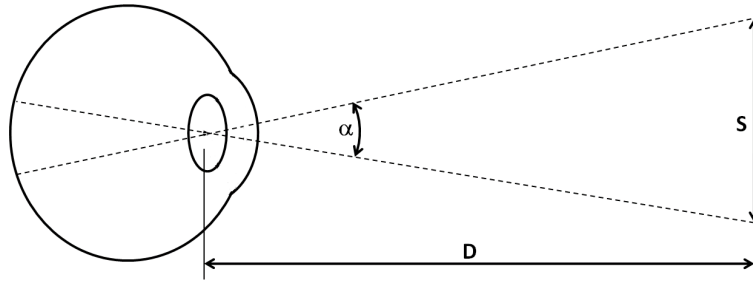


Figure 2.3: Visual angle

between lens and object and S is the object.

$$\alpha = 2 \arctan \frac{S}{2D} \quad (2.1)$$

The visual angle subtended by the fovea is estimated to be between 1 and 2 degrees, see Figure 2.4. Moving out from the fovea across the retina is the parafoveal area with around 2 to 5 degrees of visual angle. Finally, the remainder of the retina containing the fewest number of photoreceptor cells is called the peripheral area.

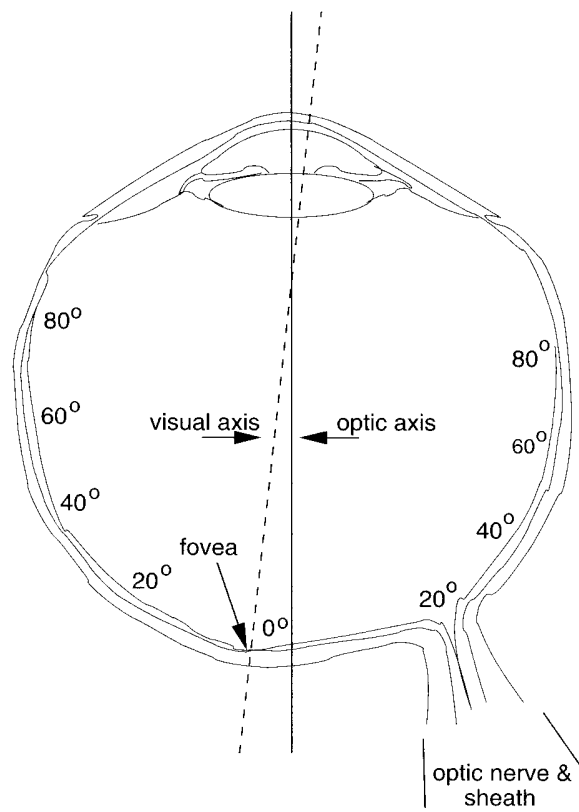


Figure 2.4: Visual angle subtended by the fovea (Duchowski, 2003).

Visual acuity is the term used to describe an individual's acuteness or sharpness of vision and usually decreases with age, with corrective lens often used to compensate. The point acuity of the fovea can be as small as several minutes of a degree for young people but drops rapidly as we move away from the fovea. The parafoveal and peripheral area are only capable of 15-50 % acuity of that of the fovea (Duchowski, 2003).

Although, point visual acuity can be measured to such fine units, the specific area to which a person is looking cannot. This is because of the 1-2 degrees of visual acuity of the fovea to which a person can still perceive sharp images. This degree of visual acuity presents an interesting problem for those researching within gaze based eye tracking as estimating where an individual's gaze is lying immediately becomes between 0 and 2 degrees inaccurate.

2.1.2 Characteristics of Eye Movements

The ability to maintain visual acuity forms the basis for the eye movements that we perform. There are three main types of movements: saccadic, smooth pursuit and vergence movement. The eye contains six muscles that control all six degrees of movement; the medial and lateral recti for left/right movements, the superior and inferior recti for up/down movements, the superior and inferior oblique for rotation movements (Davson, 1990).

Saccades are the movements made by the eye in order to reposition the fovea and fixate on different portions of a scene. They are the fastest moving external body movement, with angular movement up to 900° per second (Carpenter, 1988), traversing between 1 and 40° of visual angle at a time. Saccades are performed without feedback and once started they cannot be stopped or their trajectory altered; thus making them ballistic in nature. Once performed, a period of stability follows in which the eye is fixating upon the object.

Smooth pursuit movements are used by the eyes to visually track a moving target and can be used in combination with saccades. The initial stage of the movement is ballistic in which, the object is first observed but the visual system does not yet have the necessary visual stimuli to correct the eye position (Krauzlis and Lisberger, 1994). This stage typically lasts around 100ms after which, the process continues using a closed loop control system with negative feedback to correct the eye as the pursuit

continues. Attempts to initiate a smooth pursuit movement by tracking an object that is not moving, results in a series of small saccades.

Vergence movements are associated with obtaining binocular vision by simultaneous movement of both eyes in opposite directions. This is to ensure that the object image is projected in the centre of the retina. The eyes rotate towards each other for objects that are close (convergence) and away from each other for objects far away (divergence).

Fixations are used to describe the eye as it holds a still object within its foveal region maintaining visual acuity. During the fixation period the eye is in almost constant motion performing physiological nystagmus eye movements. These include small high frequency jittery movements of less than 1° and micro-fixations composed of slow drift and micro-saccades that are used to correct the eye when objects falls out of the foveal area during a fixation. Secondly, there are vestibulo-ocular movements or gaze-holding movements. These occur when the head is moved and the retinal image of the visual object moves across the retina. In order to compensate for the head movement the eyes move in an equal and opposite amount in the other direction. The length of a gaze-holding fixation in which an object is held within the fovea can last several seconds but this is built of a number of smaller fixations lasting between 100-300ms (Snowden et al., 2006). Further, several smaller fixations separated by saccades within a similar area are an indication of a single gaze-holding event (Just and Carpenter, 1980).

2.2 Eye and Gaze Tracking

There are two general methods of measuring eye movements: firstly, by measuring the position of the eye in head co-ordinate space (the relative motion of the eye to the head); and secondly, by measuring the gaze position in world co-ordinate space (visual line of gaze or the 'point-of-regard' (Young and Sheena, 1975)). The point-of-regard or gaze position, can be seen as a vector of infinite length starting from the fovea, going through the optical projection system into the world, until intersecting with some reference plane.

Several techniques exist for measuring eye movements and include:

- Electro-oculography (EOG)

- Scleral contact lens/search coil
- Video based (Photo-oculography/Video-oculography)

EOG and scleral contact lens/search coil are invasive methods as they require some form of attachment to the user. The latter is non-invasive when used as a remote eye tracking configuration, although there are applications for video based head mounted eye tracking.

2.2.1 Non-invasive Eye Tracking

There are several techniques used to perform POG (Photo-Oculography) / VOG (Video-Oculography) measurements depending upon what is being measured. The purpose is to take measurements of the ocular features of the eye such as, the pupil shape, the position of the limbus, as well as using corneal reflection (using a closely situated light source).

Camera based corneal reflection techniques have been used for measuring eye movements since 1901 (Robinson, 1968) and is the preferred method of measuring the gaze position. Video based eye trackers compute the gaze position in real-time with eye trackers measuring from 30hz up to speeds of 1250hz and beyond. In taking gaze position measurements, it is necessary to either fix the head so that the eye's position relative to the head and the gaze position' reference plane coincide, or by measuring ocular features to differentiate eye movement from head movement. Two common ocular features that are used are the corneal reflection and it's relationship to the pupil centre. Typically, infra-red (IR) is used as the light source and cameras fitted with IR light filters used to detect the reflection. As the light is shone into the user's eyes a number of reflections are returned from the cornea and the lens beneath. These are known as Purkinje reflections and it is the first of these four reflections, called glint, that is typically used, see Figure 2.5. The relationship between the corneal reflection and the pupil centre changes as the eye moves but barely alters upon movement of the head. This means that by simple calibration procedures eye trackers can determine the gaze position on a planar surface in which, a series of calibration points are displayed. Figure 2.6 shows the typical positioning of the first Purkinje reflection relative to the pupil when the user is looking at a series of nine calibration points positioned in segments of a screen. The reflections are typically from an IR light source which,

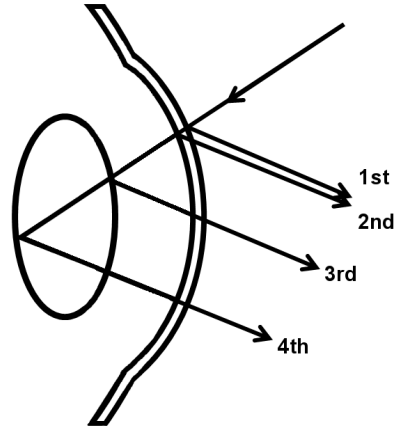


Figure 2.5: Figure showing the position of the four Purkinje reflections

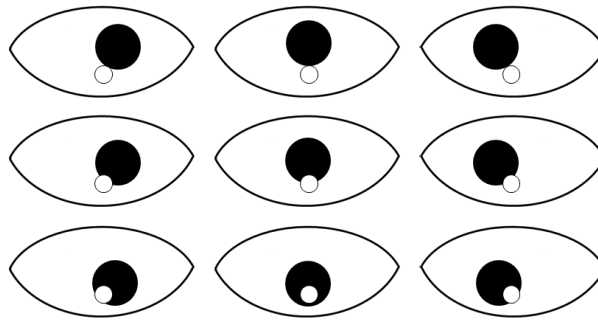


Figure 2.6: Positioning of the pupil and the first Purkinje reflection as the eye looks at nine calibration points

is invisible to the user. If the light source is located on the same axis as the camera and the eye then a bright pupil effect is created and the light is reflected back to the camera. If the light source is off axis then the camera will see a dark pupil.

2.2.2 Invasive Eye Tracking

EOG (Electro-Oculography) was a widely used method of eye movement analysis in the 1970's (Young and Sheena, 1975) however, it is only in more recent years that it has been used for interaction applications (e.g. see, Bulling et al., 2009; Gips and Olivieri, 1996; Hori et al., 2006). This method of eye tracking works by measuring the skin's electric potential differences around the eyes using a series of electrodes. EagleEyes (Gips and Olivieri, 1996) was developed at Boston College for young people

with severe disabilities to use as a communication device and uses the traditional EOG approach of individually fixing electrodes to the user, see Figure 2.7. However, the more recent developments like that of Bulling et al. have incorporated the electrodes into goggles allowing for a more compact and commercially appealing configuration.



Figure 2.7: The traditional EOG method (Gips and Olivieri, 1996) (left); EOG incorporated into a pair of glasses (Bulling et al., 2009) (right)

Scleral Contact Lens/Search Coil involves the use of an over-sized scleral contact lens that is fitted over the cornea. Attached to the lens can be optical or magnetic measuring devices which, can provide highly accurate eye movement data. Due to the nature of this method it requires much care and practice with use and is not suitable for everyday interactive applications.

Less invasive than EOG and scleral contact lens/search coil eye tracking but still requiring the user to wear/attach something are head mounted video based eye tracking systems. These eye trackers are useful in applications that are not based on human-computer interaction, where a user is free to walk around a room for example. The camera is mounted close to the eye and so there is potential for using a lower quality image sensor than with remote video based methods.

2.2.3 Open Source Eye Tracking

There have been several open source eye tracking projects over recent years with varying degrees of success. The motivation behind such projects being that commercial systems can cost several thousand to tens of thousands of pounds, leaving many people that could benefit from eye tracking unable to access it.

The EyeWriter¹ project was developed by friends of a former graffiti artist that suffers from Amyotrophic Lateral Sclerosis (ALS) . In addition to providing step-by-step instructions on how to build a head mounted gaze tracker they also developed software that allows the user to virtual graffiti the side of buildings using a projector. Researchers at the IT University of Copenhagen have been working on developing low cost eye tracking systems for several years. Recently, they have released an open source system (San Agustin et al., 2009) that works with virtually any web cam. The current system is such that the web cam must be mounted very close to the user's eye in order to get a good image of the pupil. However, the latest development allows for remote binocular eye tracking using components that total less than 100€. Other projects include, OpenEyes (Li and Parkhurst, 2006) and Opengazer².

Most of these projects require some form of do-it-yourself (DIY) that utilise web cams, High-Definition cameras, external IR sources and so on. As such, they do not offer the same levels of stability or flexibility found in commercial systems, for instance, commercial systems also allow for Z head movement (towards and away from the eye tracker) in addition to greater levels of X, Y movement. Better quality cameras, optics, algorithms and faster computing reduces the gap between low cost and commercial systems however, at this stage such systems are difficult to configure and do not yet meet the requirements for all disabled users.

2.3 Physical Disabilities

A physical disability can be any condition that has an affect on a person's body movement or control. There are many different types and they will all affect sufferers in different ways and levels. The cause of these conditions may be: hereditary or genetic; a problem encountered during birth; an illness affecting the brain, nerves or muscles; and a spinal or brain injury.

Some conditions, such as muscular dystrophy (MD) are neuromuscular and progressive (Bushby and Anderson, 2001). In the case of MD, this means that over time muscle fibres in the body will gradually weaken reducing the amount of body movement. There are many types of MD with the most common being Duchenne Muscular

¹EyeWriter: www.eyewriter.org

²Opengazer: www.inference.phy.cam.ac.uk/opengazer

Dystrophy (DMD) which, is found mostly in boys. Symptoms for DMD include, difficulty with gait and gross motor skills (walking, running etc.) with the eventual loss in the ability to walk by the age of 12 (typically); deformities in the skeleton; a high risk of neuro-behavioral (e.g. Attention Deficit Hyperactivity Disorder, Autistic-Spectrum) and learning disorders (e.g. dyslexia). DMD eventually affects all voluntary muscles, including those used by the heart and for breathing with life expectancy rarely reaching 30 (MDA, 2010).

Other conditions, such as cerebral palsy (CP) are neurological and non-progressive (Rosenbaum et al., 2007). As with MD there are many variants of CP but they are all chronic motor conditions that affect body posture, control and movement. All CP sufferers will have some difficulty with these traits but the amount will vary from mild to severe. In the more mild cases a person may have a mild limp or discomfort when walking, where in more severe cases a person may have no voluntary control over their arms, legs or even tongue. With these cases, sufferers are likely to have little or no verbal communication and can also suffer from hearing and vision difficulties. Although CP is not degenerative, existing symptoms can ease and new symptoms may appear over time. These can include involuntary movements, muscle tightness and spasm, disturbance in gait and mobility, difficulty with fine motor control, balance and walking difficulties. These symptoms contrast greatly to other neurological conditions such as Motor Neurone Disease (MND) or Amyotrophic Lateral Sclerosis (ALS), and locked-in syndromes caused by strokes and traumatic brain injuries. Here, the sufferer may be completely paralysed whilst retaining almost all cognitive function.

2.3.1 Augmentative and Alternative Communication (AAC)

People with severe physical disabilities are likely to also have difficulty in verbal communication. Augmentative and Alternative Communication (AAC) is the term used to describe the range of techniques that are used to complement or replace spoken communication. Unaided AAC, is all of the techniques that do not require additional equipment such as, vocalisation, body gestures, sign language and so on. The advantage with these is that the person will always have their communication tools with them however, not everyone will be able to understand them. In contrast, aided AAC include the techniques that require additional equipment such as, battery-free low-tech solu-



Figure 2.8: An Etran frame is a two person low-cost eye-gaze communication device. The disabled person points at groups of letters or symbols with their eyes and the other person interprets their gaze direction speaking out the letters to confirm. By courtesy of COGAIN.

tions like symbol boards to high-tech switches and eye tracking systems. Here, there are possibilities of a wider audience for understanding and greater levels of communication.

Low-Tech Aided AAC: Etran Frames

Etran frames are large sheets of clear perspex to which, symbols or groups of letters are attached, see Figure 2.8. The communication partner holds the board in between themselves and the disabled user and the user then gazes at the symbols, letters or pictures that they wish to communicate. In the Figure, the Etran user is able to communicate by selecting groups of letters, different colours and yes/no responses. This system is essentially a low-cost eye-gaze communication device and those that are able to communicate this way are likely to also be able to use an eye-gaze tracking system.



Figure 2.9: Dynamic screen communication devices.

High-Tech Aided AAC: Dynamic Screen Voice Output Communication Aids (VOCA)

Dynamic screen communication devices display groups of symbols, graphics and text on screen that can be selected in several different ways, see Figure 2.9. Multiple pages can be stored and the user is able to navigate between them as and when required. For instance, one page may give quick access to vocabulary relating to food where another page may show an entire keyboard with which, the user can type anything they like. User's can then choose to have their selections simply displayed or spoken by the device.

Selection can take place by directly touching the screen but is done mostly by using external switch selection. Those that can use two switches simultaneously will use one switch to tab between options and then use the other for selection. People that can use only a single switch must use some form of timed automatic tab scrolling and then use the switch to acknowledge selection in between tabs. Switches have the advantage in that they can be placed and positioned anywhere to best suit an individual user. Additionally, they can be of different shapes, sizes and activated in different ways such as, physically hitting, eye blinking or winking, muscle movement and so on (Figure 2.10).

2.4 Implications for Eye and Gaze Tracking

One of the deliverables by the EU funded COGAIN (Communication by Gaze Interaction) Network of Excellence was a user requirements document in which issues regarding usability and accessibility are discussed through a series of case studies (Donegan et al., 2005). Although, there is a phenomenalist approach in that no generalisation is made



Figure 2.10: A range of switches used for communication.

about disability, it does consider the implications that affect eye tracking in a general manner. These include how the eye is constructed and operates, the technology available to measure its movements, and the ability of a person to remain in the appropriate position.

Eye anatomy

Firstly there is the fovea, here we have up to 2 degrees of visual angle where clear visual acuity is obtained. This immediately creates an inaccuracy when calculating the gaze position. There are also conditions that have an affect on the structure of the eye and its ability to obtain acuity. Those who wear glasses or contact lenses may experience problems with the frames of glasses obscuring the eye trackers view of the pupil and extra reflections from the glass being caused. Contact lenses can also produce an extra purkinje-like reflection which is used rather than the reflection from the cornea. The eyelids will hide the pupil from the eye tracker during blinking and individuals that suffer from drooping eyelids may find that part of their pupil remains constantly hidden from the camera making eye tracking difficult.

Eye movements

Secondly are the movements that the eye makes in order to maintain visual acuity. Here, this can be considered to be a collection of jittery fixations connected by rapidly moving saccades. For instance, an individual is required to move their eyes up, down, left and right in order to view all areas of the computer screen. Additionally, they

2.4 Implications for Eye and Gaze Tracking

must be able to hold their gaze steady in order to fixate upon objects on the screen. Although, many people with severe physical disabilities maintain good control of their eye movements there are conditions that can have an affect on another's ability.

Nystagmus is the constant involuntary movement of the eyes, typically from side to side (RCO and RNIB, 2010) but can also be up and down and even in a circular movement. Sufferers may have difficulties fixating upon objects making dwell selection challenging.

Strabismus is a visual difficulty in which a person is unable to align both eyes simultaneously due to some form of muscle imbalance in one eye. One or both eyes may turn in, out, up or down either constantly, intermittently or alternately from one eye to the other.

Gaze paresis/palsy affects a persons ability to move both their eyes in a single horizontal or vertical direction. Strokes are a common cause to horizontal gaze paresis in which, severe cases result in no voluntary or involuntary horizontal movement and in milder cases resulting in nystagmus (Renzi et al., 1982).

Technology

Thirdly, we have the technology. For invasive methods such as EOG, the individual must have often uncomfortable eye tracking components attached to their head and body. Where as, remote video based methods limit the head movement that a user is allowed to make by keeping within the extents necessary for maintaining the pupil of the eye within view of the camera. Due to the nature of gaze tracking, there is a reliance upon the eye tracker camera gaining and maintaining a good image of the user's eye. This requires the user to remain as still as possible with the amount of head movement allowed dependant upon the eye tracker. The amount of head movement can range from no allowable movement to several inches (known as 'head box').

Body Movement

Users with CP, MD and so on can find holding their body still just as difficult as making deliberate body movements. There is a belief that people can be taught to relax their muscles thus remaining still and previous studies have shown that users can take anywhere from 15 minutes to several months to become proficient in using an eye tracker (Gips and Olivieri, 1996). However, keeping young children motivated



Figure 2.11: A young cerebral palsy sufferer sitting in his natural divergent position.

throughout this time in order to get them to a level with which, they can use an eye tracker comfortably is a challenge (Hornof and Cavender, 2005). Although, the technological improvements have eased the burden with an increased tolerance to head movements.

Many eye tracking manufacturers provide accessories for better positioning of the eye tracker and computer screen and this is particular useful for those who sit in a divergent (sideways) position. For those, who do sit in a divergent position maintaining both eyes in view of the eye tracker is particularly difficult as it is often the case that one eye may become obscured by the head rest of the seat, see Figure 2.11.

One solution to help with this is to use an eye tracker that tracks both eyes and subsequently having the software decide or switch between single eye and dual eye tracking.

Impact on Learning

There is a close association with severe physical disabilities and the impact on development and learning. Children with no physical disability usually establish good rates of balance and coordination between the ages of 7 and 12 (Berk, 2009; Brotherson, 2009). During this time schooling gradually reduce their efforts from physical improvements towards intellectual learning and growth. Children are encouraged to concentrate more

on academics and less on play, however those with a physical disability spend this period accepting a reduced level of physical ability. During these developmental years they may not be able to explore their environment in the way a child with no physical disability is able to. This lack of stimulation sets back the intellectual development and also increases levels of frustration and decreases motivation. In addition to reduced levels of communication, accessing education becomes increasingly difficult as the child gets older.

2.5 Summary

This Chapter first introduced eye anatomy and the different categories of eye movements that are made. Understanding the structure and movements of the eye is essential in understanding the technology behind eye and gaze tracking. Here, the technology has been considered to be either invasive or non-invasive, with several methods of eye and gaze tracking for each. The Chapter then goes on to discuss physical disabilities in a broad context. To cover all aspects would go beyond the scope of this thesis but it is important to gauge an understanding into some of the key issues and how they impact eye and gaze tracking. This sets the scene for the following Chapter in which, important and relevant literature on gaze based human-computer interaction and adaptive interfaces are reviewed and discussed.

Chapter 3

Gaze Based Interaction with Desktop Environments

This Chapter reviews and discusses different methods of gaze interaction used in 2D desktop and gaming environments. It begins by explaining some fundamentals of gaze interaction before discussing the different interaction techniques by application. The Chapter ends with a review on adaptive interfaces for physically disabled users.

3.1 Gaze Interaction Techniques

Eye movements are both fast as an input device and are natural as a means of pointing when compared to other input devices. Users will typically look at the area of the screen where they wish to move to before they physically operate a mouse (Ware and Mikaelian, 1987). However, a person's gaze is easily distracted by objects in the peripheral vision resulting in unwanted eye movements away from the object of interest. People don't necessarily think about the eye movements that they are making and so they are often performed subconsciously. But with practice, it is possible for a person to control their gaze such that it can be used as an effective computer input pointing device.

3.1.1 Fundamentals of Gaze Interaction

Non-command based gaze interaction

Non-command based interaction or the performing of implicit commands is where the user is merely looking around the screen and interacting without purposely trying to initiate a command (Jacob, 1993; Nielsen, 1993). This is in contrast to command based interaction in which commands are performed explicitly, typically by some kind of unnatural eye movement. There is no ambiguity in command based interfaces and it is easy for the system to determine what it is that the user requires. A non-command based interface is much more subtle in the way that it operates. The intention is that a user does not issue specific commands or instructions and the system passively observes the user and provides the appropriate responses to a user's actions.

Gaze Pointing

One approach to using gaze interaction for computer control is by gaze pointing to emulate a mouse. Here the system cursor is placed at the gaze position with the underlying application being unaware that the cursor movement, button and keystroke events are originating from an eye tracker. Stampe and Reingold (1995) found that user's who are already familiar with the desktop mouse found gaze pointing easy, intuitive and required virtually no training. In contrast however, a person with severe physical disabilities who is not familiar may take much longer (Donegan et al., 2006).

Gaze pointing allows us to place the system cursor on objects that we wish to interact with. When it comes to interacting with the object such as, initiating a mouse click, one of two methods can be used. The first is with multi-modal selection in which the user is physically capable of using a second input device, such as a switch, sensor or voice (e.g. see, Barreto et al., 2000; Fono and Vertegaal, 2005; Kumar et al., 2007; San Agustin et al., 2008; Ware and Mikaelian, 1987). The second is with mono-modal selection in which the user is unable to use a second input device and so selection must be achieved using only gaze input (examples later in this Chapter).

Multi-modal selection is achieved by the user moving the system cursor over the object that they wish to interact with and then activating their second input device. If the operation of the second input device is likely to have an affect on detecting the

user's gaze position then filtering of the gaze data is required in order to stabilise the cursor (Jacob, 1990; Stampe and Reingold, 1995).

Mono-modal selection is achieved by the user performing either a natural or deliberate eye movement in order to initiate a command. In this context, deliberate eye movements are those where the user must use their eyes in an unnatural way such as, staring at an object for a prolonged period of time or perhaps using the eyes like a pen to draw specific patterns.

Dwell

Dwell selection is the deliberate fixating or dwelling on the object of interest for a set minimum time period; a 'dwell click' (Jacob, 1990). This is usually longer than a typical fixation with times ranging from 600ms to 1500ms (Hansen et al., 2003; Majaranta and Raiha, 2002). Most eye tracking systems allow for some kind of adjustment of what constitutes a dwell. In addition to being able to vary the length of a dwell it can also be made up of several smaller fixations and so the eye tracker needs to have a position tolerance of which consecutive fixations must fall within. This often presents a problem, as it can be difficult for a person to maintain their gaze within a small specified point for the length of a required dwell. One of the disadvantages with dwell selection is the inadvertent selection of objects by looking at them for too long. This is known as the Midas Touch problem (Jacob, 1990). It is possible to set an overly long dwell time (several seconds) to partially overcome the problem but this means that task times become longer and users can become frustrated and tired (Majaranta et al., 2006). In general, there is a balance and trade off between long dwell times (longer task times, frustration and tiredness) and short dwell times (shorter task times but inadvertent selections)

The following Sections of this Chapter review the range of gaze interaction techniques based upon their application using the taxonomy found in Table 3.2.

3.1.2 Application Interface Control

This Subsection explores the different interaction techniques that have been applied to controlling the application interface, for instance interacting with dialog boxes, small

3.1 Gaze Interaction Techniques

<i>Application</i>	<i>Interaction Technique</i>	<i>Reference</i>
<i>Application interface control</i> (Subsection 3.1.2)	Gestures	Drewes and Schmidt, 2007
	Gaze pointing	Zhai et al., 1999
	Zooming	Istance et al., 1996; Lankford, 2000; Skovsgaard et al., 2008
<i>Text entry & communication</i> (Subsection 3.1.3)	Gestures	Bee and Andre, 2008; Wobbrock et al., 2008
	Zooming	Hansen et al., 2008; Ward et al., 2000
	Dwell keyboard	Istance et al., 1996; Johansen et al., 2003; Hansen et al., 2003; Majaranta, 2009
	Pie menu	Huckauf et al., 2005
<i>Drawing</i> (Subsection 3.1.4)	Gestures	Tchalenko, 2001; Heikkila and Raiha, 2009
	Gaze pointing	Gips and Olivieri, 1996; Hornof and Caverder, 2005
<i>Reading & Internet browsing</i> (Subsection 3.1.5)	Dwell buttons	Castellina and Corno, 2007; Kiyohiko et al., 2007
	Non-command	Kumar and Winograd, 2007
<i>Gaming</i> (Subsection 3.1.6)	Gestures	Møllenbach et al., 2010 Istance et al., 2010
	Gaze pointing	Isokoski et al., 2009; Jönsson, 2005
	Mouse emulation	Smith and Graham, 2006
	Non-command	Starker and Bolt, 1990; Dorr et al., 2007

Table 3.1: A taxonomy of gaze interaction technique applications for human computer interaction. The references are not an exhaustive compilation, merely a sample.

icons and so on. The typical technique used to perform this task is mouse emulation with dwell selection. Two alternative techniques are reviewed: zooming and gaze gestures.

Zooming techniques are of interest for gaze interaction as they potentially account for the noise and inaccuracies that are often found in gaze tracking. A zooming interface allows the user to change the scale of the viewed area so that they can see more or less detail. Istance et al. (1996) allowed users to magnify parts of the desktop by dwelling on the area of interest. The magnified area would appear elsewhere on the screen and the user is able to use mouse emulation and then dwelling on the magnified controls. Here, the zoomed area was part of an on-screen keyboard and remained on-screen for the entire time. A similar zoom method was also implemented by Lankford (2000) with the exception that the zoomed area only became visible when the first dwell occurred. Both implementations allowed the user to change the amount of zoom.

Skovsgaard et al. (2008) (and Skovsgaard et al. (2010)) experimented with several methods of zooming for selection of small desktop controls of 6 and 12 pixels in diameter. Similar to Istance et al. and Lankford they used a ‘two-step dwell’ method of zooming in on the control. They also used a dynamic continuous zoom method in which, the region surrounding the cursor would increase in size as a user started to fixate within an area. When the zoom reached a certain level, a selection would be made at the cursor. The expectation was that this kind of dynamic zoom would perform more naturally and make selection faster than using a two-step dwell select method however, they found that using the dynamic zoom method changed a selection task into a continuous tracking task. This is because as the zoom was occurring the control object would move relative to the gaze/cursor position and so would require the user to track the moving object until the selection was made.

Gaze gestures are sequences or patterns of eye movements that are interpreted by the eye tracker software and used to initiate commands. Istance et al. (2010) defines a gaze gesture as:

‘A definable pattern of eye movements performed within a limited time period, which may or may not be constrained to a particular range or area, which can be identified in real-time, and used to signify a particular command or intent.’

They are a familiar concept with other input devices such as, stylus, mouse and body and these sources have provided the inspiration for several gaze gesture implementations. Drewes and Schmidt (2007) developed a gesture recogniser inspired by Wob-

brock's EdgeWrite (see, Wobbrock et al., 2003) and the Firefox web browser mouse gesture plug-in¹. Gestures can be made up using combinations of eight different eye movement directions as in Figure 3.1, with three of the gestures used in their experiments in Figure 3.2.

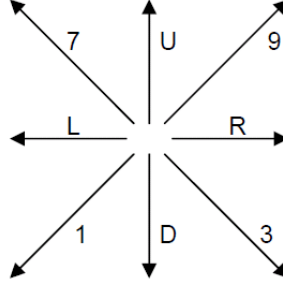


Figure 3.1: The eight different eye movement directions used in the gesture recogniser by Drewes and Schmidt

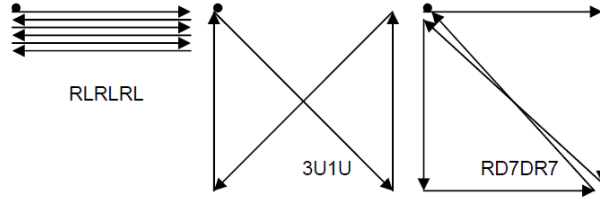


Figure 3.2: Three of the gaze gestures used in Drewes and Schmidt user study

The authors conducted two experiments, the first in some basic interface control and the second on performing gestures on different backgrounds. In their first experiment participants were required to close a dialog box by looking around the corners of the box in a clockwise direction. The gesture recogniser would be constantly scanning for, using their notation: RDLU, DLUR, LURD or URDL, gesture patterns and when one occurred the system would close the dialog box. The experiment also requires participants to perform the gesture in the other direction with the resulting action being that the dialog box remained open. They found that all participants were able to complete the task but with a mean time of 1905ms (dialog closing) and 1818ms (dialog remaining open). In the second experiment, nine participants were required to perform the gestures in Figure 3.2 over three different backgrounds: with guidelines

¹<http://optimoz.mozdev.org/gestures/>

to help the gesture; with a text background; with a blank background. They found that the background didn't affect the persons ability to perform a gesture and complex gestures can still be performed over a structured busy background such as text or tables. The time to complete each gesture was dependent upon the number of strokes that it contained, with the mean time for one stroke being 557 ms.

Although, not a gaze only interaction technique, Zhai et al. (1999) used gaze position to enhance the speed of desktop mouse interaction. Manual And Gaze Input Cascaded (MAGIC) Pointing is based upon the idea that a person looks at the object that they wish to interact with before they interact with it. This may involve moving the mouse cursor from one side of the screen to the other. In the authors implementation, the cursor first jumps to the gaze position when it is within the vicinity of the object. The user can then fine-tune the cursor position by moving the mouse. One problem with this approach however, is that there are many occasions where a user wants to look at parts of the screen without interacting with it for instance, reading passages of text. With this implementation the cursor would always (eventually) arrive at the gaze position, and without a method to turn it off it is likely to distract the user from the task that they are performing.

3.1.3 Text Entry and Communication

This Subsection explores the different interaction techniques that have been applied to text entry or eye typing. In addition to the traditional on-screen keyboard method of eye typing; gaze gesture and zooming methods are also reviewed.

Eye typing using an on-screen dwell keyboard is a well researched area with systems available since the late 1970s (Majaranta and Raiha, 2002). With such keyboards, it is not necessary to adhere to an ordinary QWERTY key layout and many allow for the layout to be changed, see Figure 3.3. This is important as there is a balance between the number of keys, the size of the keys and the amount of screen space taken by the keyboard. The smaller the keys are then the harder they become to select; the closer they are together the higher the risk of accuracy problems causing incorrect selections. Some virtual keyboards can dynamically change keyboard content and offer word and sentence prediction offering an increase in typing speed (e.g. see Figure 3.4). This also allows for the possibility of dynamic dwell time selection by reducing the length of time required to select keys that are more likely required and increasing the time



Figure 3.3: A virtual on-screen keyboard used for gaze typing by Istance et al. (1996)

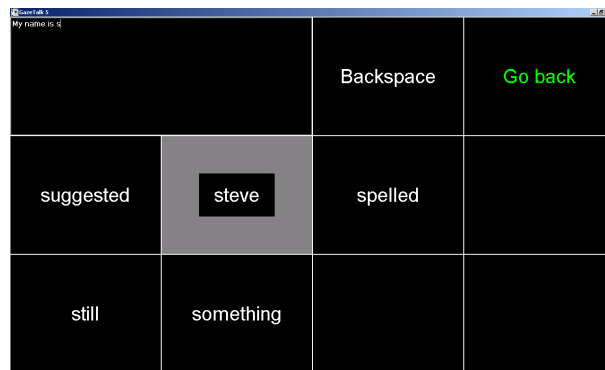


Figure 3.4: GazeTalk's full screen keyboard with large buttons that are easy to select and word prediction to increase text entry speed. GazeTalk has been particularly successful in that it is one of the few examples of a research application that is used on a daily basis by people with severe physical disabilities

for those are not. This method allows for a reduction in dwell time by 20-65% thus increasing overall typing speed (Nantais et al., 1994; Salvucci, 1999). The ability to increase typing speed is particularly advantageous when eye typing is also being used as a means of communication (via synthesised speech, for instance GazeTalk (Johansen et al., 2003)).

When a user types with an ordinary desktop keyboard they receive haptic feedback from the keys as they are pressed. This sensation along with the letter's visually appearing on the screen is part of the user's acknowledgement that they have pressed the key successfully. When eye typing the user's focus is on the key that they wish to select, so they first require feedback that their gaze position is on the correct key, then once the dwell time period has expired they require feedback to confirm selection (Majaranta and Raiha, 2002). Istance et al. found that user's would make mistakes if not given adequate feedback thus leading to errors and subsequent correction of

text, resulting in a much longer text input speed. After the key has been selected the associated letter is presented somewhere on the screen. This can be close to where the virtual keys are or elsewhere on the screen. Instance et al. also observed that users were either checking letters entered one at a time (resulting in slower but more accurate text entry) or inserting several letters before checking them (resulting in quicker but less accurate text entry; if a mistake occurred then correction of several letters is more complicated than correcting one) .

Bee and Andre (2008) implemented a gaze variant of Quikwriting¹ (see, Perlin, 1998). The idea had been proposed previously by Isokoski (2000) although, was never built by him. In Perlin original version the user can search the regions for the character that they wish to write and then perform the corresponding stylus gesture. As it is not possible to search the regions with gaze without a gesture being performed, Bee and Andre placed the characters within the central region, see Figure 3.5. The user is then

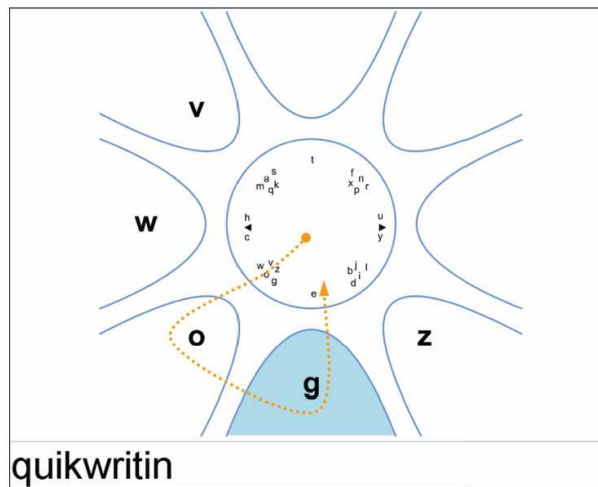


Figure 3.5: The Quikwriting interface adapted for use with eye gaze. The dotted line shows the gaze path required to write the character ‘g’

able to plan which regions they need to complete the gesture without leaving the safety of the central region. The location of the typed text also causes difficulties when using gaze. When inputting text via gaze, the user will often look for feedback checking that the letter that they have just entered is correct (Instance et al., 1996; Majaranta and Raiha, 2002). In this instance the user’s gaze position must pass through the gesture

¹Quikwriting is a stylus based text entry application.

zones meaning there is a possibility of generating accidental gestures. This problem was partially overcome by the system disabling writing actions whenever the user's gaze position was in the text field. Although, during experiments they found that this did not work successfully and resulted in many errors occurring. When compared to a dwell activated on-screen keyboard (750 ms) they found that participants were able to achieve 5.0 words per minute (WPM) using the Quikwriting method in contrast to 7.8 WPM from the keyboard. Wobbrock et al. (2008) implemented a similar system called EyeWrite that was based on the EdgeWrite text entry application that they had previously developed (Wobbrock et al., 2003). Users are able to enter characters by drawing letter like gesture sequences with their eyes. The letters are based on a neography of upper-case EdgeWrite letters, see Figure 3.6, and users draw the letters using the separated corners of the application interface, see Figure 3.7. The letters are similar to those found in Palm Inc. Graffiti applications, and are designed to be easy to learn as they are similar in shape to the letters that they represent.

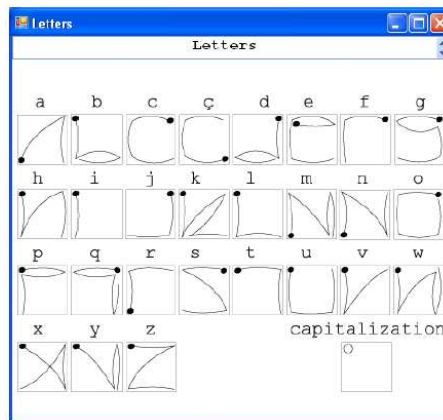


Figure 3.6: EyeWrite's letter chart

Once a letter gesture sequence has been made the user must finally look (and fixate) at the dot at the centre of the window sending the character to the target application. They found that their system was slower for text entry than a dwell (330 ms) on-screen keyboard (EyeWrite mean = 2.5 WPM; on-screen keyboard mean = 6.14 WPM) but projected learning curves suggest that the speed could approach that of an on-screen keyboard. Although, the perceived speed and ease of use was considered to be worse with EyeWrite; the fatigue experienced through performing this kind of gesture was

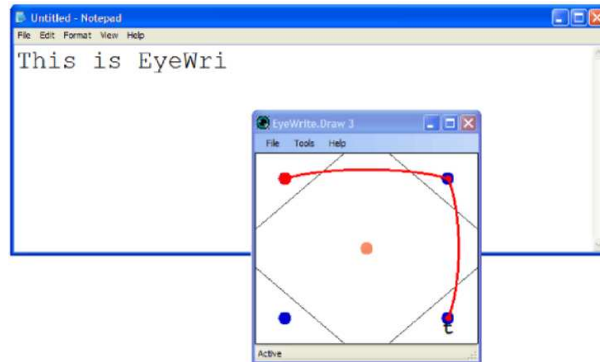


Figure 3.7: EyeWrite being used with Notepad. The current sequence shows a letter ‘t’ being created.

much less than dwelling. Possible reasons for such a large difference in performance between Quikwriting and EyeWrite maybe due to EyeWrite’s requirement of having to learn all of the letter gestures; the additional dwell that is required at the end of every letter; the extra gesture stroke that is sometimes required.

Hansen et al. (2008) used dynamic continuous zoom as a method for text entry with StarGazer. The user is presented with a set of concentric circles in the centre of the interface surrounded by a circle of letters. As the user looks around at the letters, the three circles appear to move, zoom and point towards the gaze position giving the appearance of flying in 3D towards them. The three circles always stay in the centre of the interface and so the letters in fact pan across to the centre as the user looks at them. The user selects letters by flying through them after which point, the view is reset to its original state, see Figure 3.8. An advantage of such a zooming interface is that there are possibilities to make selections that appear context sensitive. A lot of higher level information can be presented on screen and as the user flies closer to a particular area then the information is broken down into smaller pieces. The authors used this to present the circle of letters in different ways in addition to offering word and letter prediction.

Although not specifically designed for eye-gaze, Dasher (Ward et al., 2000) allows user’s to insert text by pointing at letters as they enlarge and fly horizontally across the screen, see Figure 3.9. It has been used with a manner of pointing devices of which, eye-gaze has been particularly successful (Ward and Mikaelian, 2002). Letters start at the right hand side of the screen and as the user moves their gaze towards them they

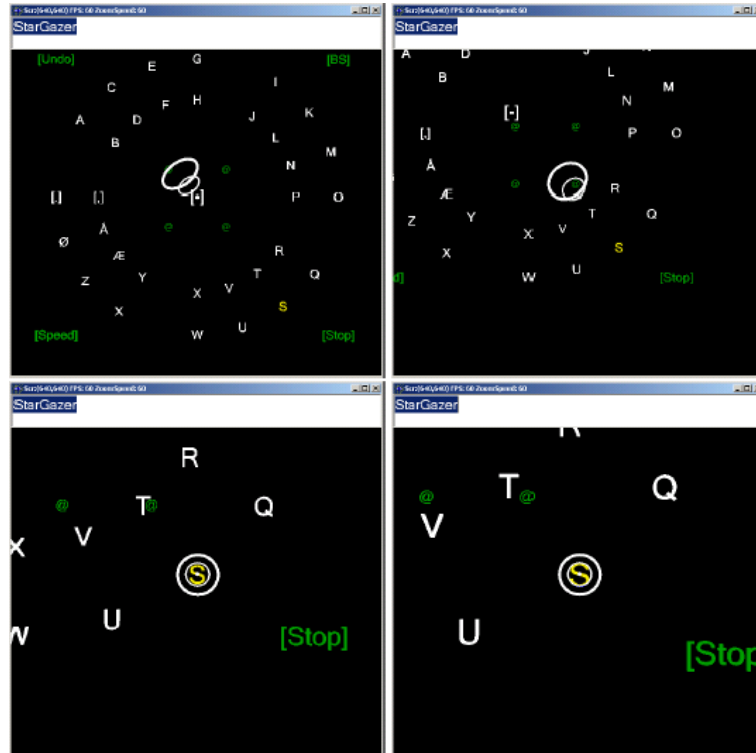


Figure 3.8: Using StarGazer to select the letter ‘S’

begin to grow in size slightly and then start to move to the left and across the screen. During this time, the systems language model makes predictions on what the next letters are likely to be. These predictions appear as new letters behind the previous letter and as such, closer to where the user is pointing, thus reducing the distance and time to select the next letter. As the letters cross a vertical line in the centre of the screen the letters are selected. To delete an incorrect letter the user has to just look to the left of the vertical line and the letters begin to fly back across to the right. As the letters pass back over the vertical line they are deleted. Such an interaction technique allows for free continuous operation without the need for dwells or specific gestures.

3.1.4 Drawing

This subsection explores the different interaction techniques that have been applied to drawing or eye painting applications.

Eye painting can be achieved by using existing desktop applications such as Microsoft’s Paint and gaze based mouse emulation, with the user using dwell to select the

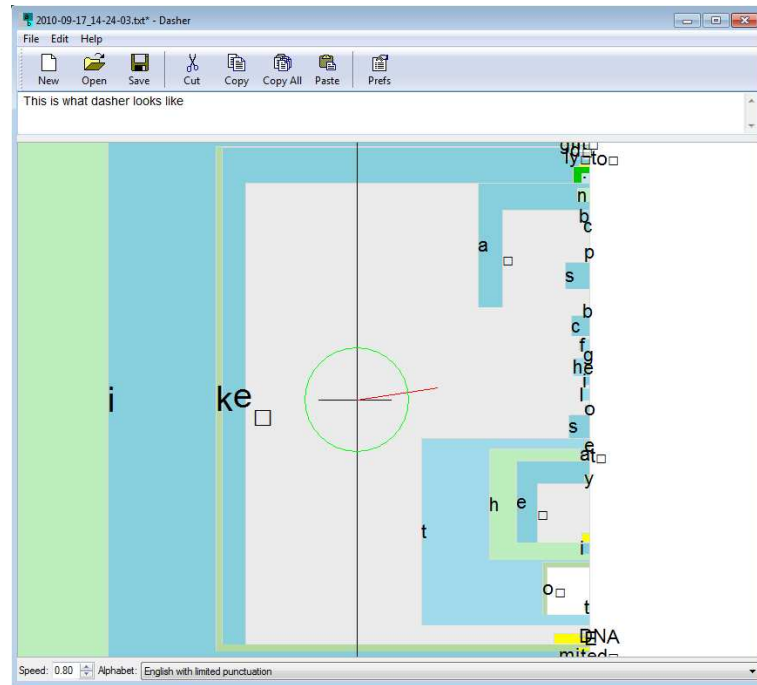


Figure 3.9: Dasher’s interface that uses an interaction technique where text is entered by flying into letters

pen and colour and then having the pen ink follow the user’s gaze path around the drawing area. Straight lines can be drawn easily by simply fixating at the start of the line and then again at the end. However, drawing curves is more difficult as the eye does not move smoothly in this context but in short saccadic jumps (Tchalenko, 2001). Gips and Olivieri (1996) EaglePaint painting application, used a free eye-drawing style of gaze finger-painting. Here, ink would follow the gaze path of the user with the ink appearing to change colour after each fixation, see Figure 3.10. As this is free eye drawing, the ink is in an always-on mode with no way for the user to disengage it. When the user has finished painting or looks away from the screen then the ink will follow them to the end of the screen creating an unwanted final line. Hornof and Cavender (2005) used a different approach with EyeDraw by allowing the user to switch between looking and drawing by performing a short dwell (500ms). Thus allowing the user to efficiently start and stop drawing as they wish, see Figure 3.11 for an example. However, they found that the younger children in their study (9 yrs and 12 yrs) had difficulties in learning how to use the application and so suggest that free eye-drawing methods like



Figure 3.10: Finger painting inspired eye drawing (Gips and Olivieri, 1996)

that of EaglePaint should be used as a starting point.

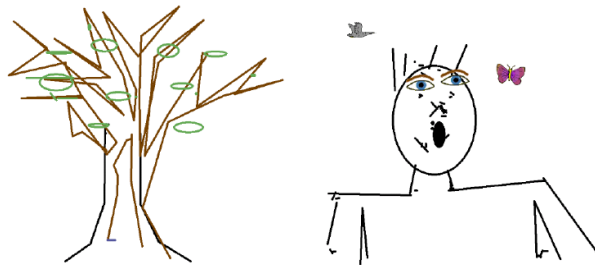


Figure 3.11: Drawing made by a child with EyeDraw (Hornof and Cavender, 2005)

Although not used directly for eye painting, Heikkila and Raiha (2009) investigated the accuracy of gaze gestures and what affect the size of the gesture has on speed and accuracy through a series of drawing based tasks. The first series of tasks were to determine the ability in drawing an instructed shape and required the participant to draw: the letter ‘L’, a triangle, and a horizontal line, with their gaze on an empty drawing area. The second series of tasks was to determine the ability to follow a specific model and required the participants to draw a rectangle, a circle, and a horizontal line, with their gaze on a drawing area that had the model shape. The participant was not given any feedback on their gaze position. Their results showed that the mean time taken per stroke varied from 824ms (when drawing the letter ‘L’ on a plain background) to 1190ms (when drawing a rectangle with guidance model). Interestingly, they found that when gestures were performed on a larger scale they did not take longer to perform. Thus, the distance in which a saccade must travel has little affect on the overall gesture

time. A conclusion that was also reached by Drewes and Schmidt. In terms of accuracy, Tchalenko had previously found that although drawing curves was possible with gaze and is achievable with some accuracy however, it is a difficult task as it requires many short saccades. This was also found to be true with Heikkila and Raiha in which the circle was drawn accurately but consumed a lot of time and concentration.

3.1.5 Reading and Internet Browsing

This Subsection explores the different interaction techniques that have been applied to on-screen reading and internet browsing. This includes dwell button techniques and techniques designed specifically to be non-command based.

There are a number of commercially available internet browsing applications¹ developed specifically for use with eye-gaze. Typically, the application will scan the desired web page for controls and objects such as hyper-links, radio buttons, text boxes and so on and then allow the user to easily move from one to the next. Movement through the objects and scrolling of the web page being achieved using dwell activated buttons located outside the browser. This is a significantly easier approach than trying to navigate a web page using a standard web browser and gaze based mouse emulation. Whilst Kiyohiko et al. (2007) used a modified web browser interface operated by gaze, Castellina and Corno (2007) approach was to develop an add-on for an existing web browser. Most modern web browsers allow for developers to extend their browsers functionality by creating add-ons. The authors Accessible Surfing Extension (ASE) runs within Mozilla Firefox as a side bar application. Using this method it is possible to add a gaze friendly interface to a stable existing application.

Jacob (1993) implemented an interaction technique that allows a user to scroll through text. A row of arrows appears below the last line of a series of text indicating that there is more text to be seen. When the user looks at the arrows the text begins to scroll and to stop scrolling the user looks away. The disadvantage of this technique is that as the user is scrolling they are no longer looking at the text and therefore have no indication on how much text has been scrolled through. This problem was slightly addressed with Kumar and Winograd (2007) who used automatic scrolling of text based on the current sentence that was being read by the user. Although, several versions were trialled, one in particular used automatic gaze enhanced scrolling that

¹For instance, Tobii Communicator: www.tobii.com; Dynavox EyeMax: www.dynavox.com

required no purpose command control by the user. Thresholds vertically positioned on the screen (Figure 3.12) were used to activate the scrolling and also change the scrolling speed. During their studies, they found that participants were happy with the speed

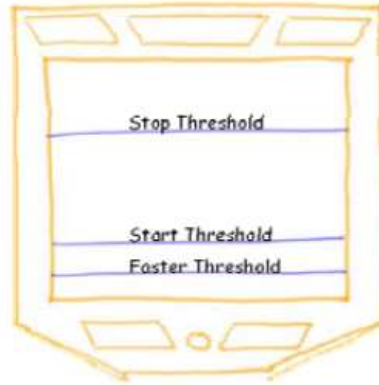


Figure 3.12: The vertical thresholds used by Kumar and Winograd to determine when automatic scrolling should take place when reading text

and change of speed of the scrolling but they found reading the text as it scrolled to be disconcerting. Although, they also reported that once familiar with how the scrolling worked users eventually relaxed and were able to read comfortably.

3.1.6 Gaming

This Subsection explores the different interaction techniques that have been applied to gaming. There have been several investigations into using eye-gaze as an additional input device for gaming but few that examine using eye gaze on it's own. The following includes gesture, eye pointing and non-command based interaction techniques.

Isokoski et al. (2009) used a First Person Shooter game in order to assess the performance of eye gaze as an extra input modality to mouse and keyboard. The control system of such games uses W,A,S,D/cursor keys for movement and mouse for camera control and shooting. The trials required subjects to walk around a bespoke game environment and shoot at moving targets as they appear. Eye gaze was used for aiming of the weapon, with an XBox 360 controller to move the avatar. Their first findings showed that eye gaze will not always improve the performance of the players when compared to using the hand controller for aiming instead of gaze, but they found that the number of hits from gaze is comparable to the game pad controller

used alone, and that gaze was often more entertaining. The possibility of using eye gaze for controlling the player direction was also briefly examined, but due to the game requiring the user to change direction very quickly and issues with jitter filtering it was not deemed feasible.

Smith and Graham (2006) performed an experiment using a similar control system (gaze to aim, mouse to move the camera, keyboard to move avatar) on an open source port of the game Quake 2¹ called Jake2². Similar to Isokoski et al., Smith and Graham did not find any advantage in performance. However, their subjective user results showed that using eye gaze offered a more immersive experience than using a mouse and keyboard. Jönsson (2005) also found that a combination of mouse and eye gaze was more immersive than mouse and keyboard when performing trials using the first person shooter Half Life³.

Smith and Graham also performed trials using a version of the 80's arcade game Missile Command. In this game the user is required to defend several cities at the bottom of the screen from missiles falling from above. The missiles are to be destroyed by the user intercepting their flight path with a gunshot. Trials were conducted using a mouse to aim and shoot, and then by using gaze to aim and a button to shoot. The trials showed that there is a need to fire far ahead of the missile path and this is easily achieved using a mouse. However, it is difficult when using eye gaze to fire ahead, as there is a constant distraction of the missile itself (the users looked at the missile rather than where they wished the missile to go). Thus, the majority of eye gaze shots missed and fell behind the missile. However, a similar style of game, Breakout, was tested by Dorr et al. (2007), and they demonstrated in their implementation that eye-gaze can outperform a mouse. Breakout involves hitting and destroying bricks by bouncing a ball off a paddle that the user controls at the bottom of the screen. The paddle is moved in a horizontal direction only and so is simply controlled with left and right movements of a mouse or cursor keys and these can be easily translated for use with gaze. The trials required users to compete on a one-on-one basis: one user with gaze and the other with mouse. The results showed that two-thirds of all rounds (65 out of 99) were won by the gaze player. The game play using gaze was so natural and effective

¹Quake 2: www.idsoftware.com/games/quake/quake2

²Jake2: www.bytonic.de/html/jake2.html

³Half Life: www.orange.half-life2.com

that one user did not realise she was even playing the game and after 2 minutes asked when the experiment would actually start. This was due to her constantly watching the ball, meaning that the paddle would follow her gaze and so ball and paddle would always intersect, provided she watched the ball.

Whilst being more of an interactive storybook rather than a game, Starker and Bolt (1990) example of a gaze responsive self disclosing display is an early example of non-command based interaction. The authors approach was to generate a model of user interest as a way of directing a story. Using the book *The Little Prince* by Antoine Exupery as a theme, users were presented with a 3D graphical environment as the story is narrated. The user is free to look where they wish but the direction of the narration is dependant upon which, objects on the screen that the user is most interested.

Although not specified as a game, Møllenbach et al. (2010) examined the use of single stroke gaze gestures in a game like experiment. Their studies required participants to observe coloured blocks falling from the top of the screen and then perform a horizontal or vertical gesture to identify their colour, see Figure 3.13. Two different interfaces were designed that allowed participants to make selections based on long gestures and short gestures. As a block falls the participant must first initiate the start of the gesture by fixating within the relevant coloured area and is then given a 1000ms time limit to fixate in the opposing coloured area. If they exceed this time period then the gesture is reset.

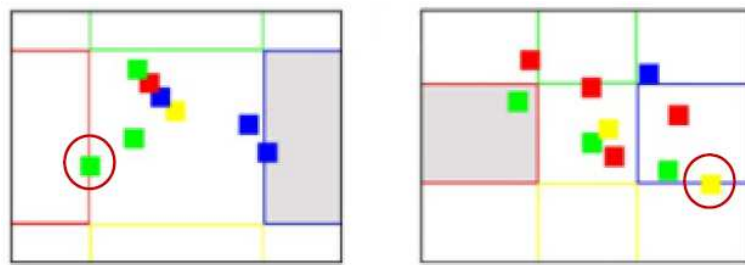


Figure 3.13: Møllenbach et al. two single stroke gaze gesture interfaces. Long gestures (left) and short gestures (right)

In contrast to Drewes and Schmidt and Heikkila and Raiha, their results found that there was a significant difference between performing short distance gestures (mean = 78.81ms) and long distance gestures (mean = 131.45ms). A suggested reason for this is that the short gestures were foveal and the long gesture peripheral. However, the

idea of a foveal gestures suggests that the saccade in between the two fixations be only a few degrees of visual angle (i.e. a gesture that is subtended within the foveal area). Additionally, they found that horizontal gestures (mean = 139.88ms) were performed significantly faster than vertical gestures (mean = 185.32ms). The length of the gestures is much shorter than those reported by others (see Table 3.2 for a summary) and one possible explanation for this is that the experiment appears similar to a game and so participants were trying to complete the task as quickly as possible.

<i>Publication</i>	<i>Performance</i>
Bee and Andre (2008)	5 WPM
Wobbrock et al. (2008)	2.5 WPM
Drewes and Schmidt (2007)	557 ms per stroke
Heikkila and Raiha (2009)	824 ms per stroke
Møllenbach et al. (2010)	78 ms / 131.45 ms per stroke

Table 3.2: Performance comparison of different gaze based gesture implementations

3.2 Adaptive Interfaces

Although, the applications and interaction techniques discussed previously are designed for use by eye-gaze they all require expert configuration so as to suit an individual. Configurations such as, changing the dwell select time; changing the size of the interface buttons and their location on screen; creating custom dwell keyboard configurations for communication; the amount and speed of a zooming interface; all require configuration and expert knowledge and so is likely to be beyond the capabilities of a parent, carer or support worker (Dawe, 2006).

Previous work, on user interface adaption for physically disabled users include, adaption of web pages (see, Bigham et al., 2006; Mankoff et al., 2002); automatic generation of GUIs (see, Gajos, 2008; Gajos et al., 2008) and dynamic adaption of GUIs (Carter et al., 2006). No specific research into interfaces automatically generated to suit an individual using eye-gaze could be found; this has previously been a manual process.

Of particular interest to this work is the research by Gajos as it is aimed at those with motor disabilities, such as cerebral palsy and muscular dystrophy. Gajos developed two systems for automatically generating user interfaces. The first system SUPPLE, uses a preference elicitation engine in order to model a preferred user interface configuration. The elicitation engine is a computer guided process with which the user must select their preference out of a pair of user interface fragments presented on screen. The fragments differ in presentation but functionally would be the same. The second system SUPPLE++, models user's motor abilities from a set of one time motor performance tests. Here, four task types are used (also see Figure 3.14):

- Pointing : A set of tasks based on ISO9241-9 standard (ISO, 2000), where the user is required to move the pointer to targets at various sizes and distances
- Dragging : A task designed to be similar to the action of using an interface scroll bar component
- List selection : A task where the user is required to make multiple selections from a given list of menu items or scroll bar components
- Multiple clicking : A task where the user is required to click within circular targets of varying sizes.

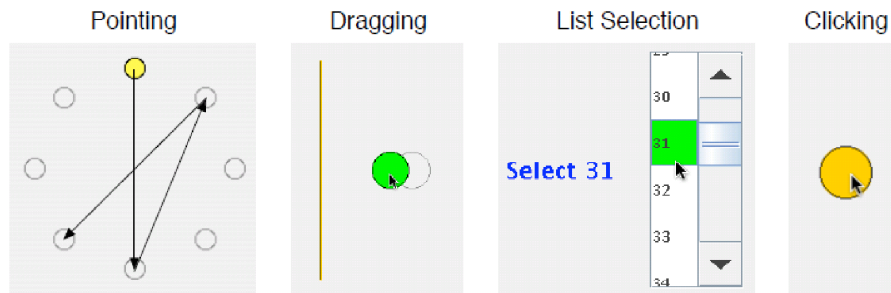


Figure 3.14: Four task types used to measure participant's motor capabilities, SUPPLE++

The type of input device used by participants in their studies was dependant upon the individual and included: mouse operated by fingers, one hand or two hands; track-ball operated by chin, backs of the fingers, back of the hand and bottom of the wrist.

The range of conditions by participants included cerebral palsy, muscular dystrophy, spinal cord injury and Parkinson's. Based on the outcomes of the models two interfaces will be generated, one based on preference and one based on ability, see Figure 3.15.



Figure 3.15: Two different user interfaces based on SUPPLE preference elicitation for two different users (AB03) and (MI09). The first interface is the baseline interface.

They compared the two interfaces along with a traditional dialog interface and measured the individuals ability (in time and subjectively) to complete a series of tasks. They found that participants were able to perform the tasks 10% faster using the preference interface and 28% faster using the ability interface than when using the base-line. Additionally, users made 73% fewer errors when using the ability interface and subjectively, strongly preferred (in terms of efficiency and ease of use) both the preference and ability interfaces.

3.3 Summary

Gaze interaction techniques are a fundamental theme within this thesis and so it is important to know the strengths and weaknesses of previous work. The techniques discussed in this Chapter were reviewed within a taxonomy based on desktop computer applications. This was to allow easy reference and also to see the scope of tasks that each technique offers. There was no research found on adapting gaze interaction techniques but research conducted on adaptive interfaces for physically disabled users is in part relevant to the work in this thesis. Although, this work is based on mouse pointing and interface customisation the idea of presenting diagnostic tests to users is considered novel and discussed later in this thesis.

The following Chapters explores the use of eye-gaze for control of 3D online games and the possibility of selecting candidate interaction techniques for specific in-game tasks.

Part II

Gaze Interaction Techniques for Game Tasks

Chapter 4

Initial Assessment of Gaze Interaction with Online Games

The aim of this Chapter is to first describe Multi-User Virtual Environments (MUVE) and Massively Multiplayer Online Games (MMOG) and examine the control demands that these environments have on disabled users. Then, after exploring the different task categories the Chapter goes on to present an initial pilot study of gaze based mouse emulation within an MUVE.

This Chapter is based on the following journal article and conference publication. Some portions are reproduced as are but much is rewritten to suit the flow of the Chapter and the thesis. All authors have made essential contributions and a brief note on my contribution personal is shown here.

Bates, R.; Vickers, S. & Istance, H. **Gaze Interaction with Virtual On-Line Communities: Levelling the Playing Field for Disabled Users.** *Universal Access in the Information Society; UAIS*. SpringerLink, 2009. 9. 261-272

Howell Istance coordinated the pilot study and Richard Bates prepared the first draft. I wrote the literature review and parts of the discussion.

Vickers, S.; Bates, R. & Istance, H. **Gazing into a Second Life: Gaze-Driven Adventures, Control Barriers, and the Need for Disability**

4.1 Multi-User Virtual Environments (MUVE) and Massively Multiplayer Online Games (MMOG)

Privacy in an Online Virtual World. *Proceedings of the 7th International Conference on Disability, Virtual Reality and Associated Technologies; ICDVRAT 2008.* 2008

This paper was based on the pilot study reported in the previous article. I wrote the full version of this paper.

4.1 Multi-User Virtual Environments (MUVE) and Massively Multiplayer Online Games (MMOG)

Online virtual environments are becoming increasingly popular as a means of interacting, chatting and spending time with friends and new acquaintances. Second Life, Free Realms, World of Warcraft and so on are part of the growing family of Massively Multiplayer Online Games (MMOG) and as computers and the internet become faster these worlds become more realistic and immersive. Second Life, for example, has over 14 million members (Linden-Labs, 2011) and can support interest groups, cooperative activities between people as well as serious commercial activities. In contrast, the Massively Multiplayer Online Role Playing Game (MMORPG) of World of Warcraft is more focused on character development achieved through completing game related goals (see Figure 4.2). World of Warcraft has more than 12 million users worldwide (Blizzard, 2010). There are also MMORPGs designed specifically for children in which, character development also involves completing educational elements such as maths, reading and language skills (e.g. Free Realms, see Figure 4.3).

Within these communities the users are represented as a virtual projection of themselves in the form of an avatar. The user can choose to appear as any shape, size, colour or other appearance that the game customisation allows. Depending upon the aim of the game, users are free to move around by walking, running or even flying. Objects can be created from simple shapes to complex virtual homes and even moving objects such as cars or spaceships. As these worlds are virtual, then almost anything creative may be possible. The three figures show a view of the world from a third person perspective, with our avatars back to us and the camera placed behind the avatar.

4.1 Multi-User Virtual Environments (MUVE) and Massively Multiplayer Online Games (MMOG)



Figure 4.1: ‘Second Life’ created by Linden Labs - <http://www.secondlife.com>. In this MUVE, there is no character development but an open world to explore or contribute to.



Figure 4.2: ‘World of Warcraft’ created by Blizzard - <http://eu.battle.net/wow/en/>. In this MMORPG, player’s develop their character through completing quests that typically involve fighting monsters.

4.1 Multi-User Virtual Environments (MUVE) and Massively Multiplayer Online Games (MMOG)



Figure 4.3: ‘Free Realms’ created by Sony - <http://www.freerealms.com>. In this MMORPG, player’s develop their character through quests that involve educational elements such as ‘herding 7 grey sheep into a pen’ or ‘collect these ingredients in order to make a pancake’ and so on.

4.2 Task Interaction

To survey interaction in MUVES and MMORPGs it is necessary to determine the main type of tasks that occur. Hand (1997) previously proposed a taxonomy of main control and manipulation areas that are present in virtual environments: locomotion and camera movement; object manipulation; application control; communication (previously not specified by Hand). Taking Second Life as an example:

- Locomotion and camera movement: These both may be controlled by using arrow buttons located on semi-transparent overlays as can be seen in Figure 4.4. Continuous motion is performed by holding the mouse button and performing a ‘dragging’ motion. There are also keyboard shortcuts to perform avatar movement by using the cursor arrow keys. In many online virtual environments there are also possibilities to use the ‘W, A, S and D’ keys to perform movements.
- Object manipulation: This is only achievable through mouse control. In order to manipulate an existing object then it is selected with a right mouse click, at which point a semi-transparent pie menu appears at the point of click, see bottom right of Figure 4.1. The pie menu offers several options related to the object such as ‘Open’, ‘Edit’ and ‘Sit Here’. A new object is created by selecting the ‘Create’ option within the pie-menu. This causes a dialog box to appear offering basic functions similar to those found in 3D modelling packages.
- Application control: This is mostly accomplished using the mouse although there are several commands and menus accessible via the keyboard. Menus must be opened using left button mouse clicks and although some commands can be accessed using a keyboard the majority of the menu functions are available using mouse only. One of the major functions within Second Life is the changing of the avatars appearance and is only accessible through using a mouse, apart from keyboard tabbing between buttons and using arrow keys with slider controls.
- Communication: This is achieved through text generation or speech relay via a microphone. A chat box lies at the bottom left of the screen, see Figure 4.1, and text generated through conversation is displayed. Speech relay allows the voice of the user to be heard by nearby avatars, together with hearing any nearby avatars.

4.3 Pilot Study 1 - Initial Assessment of Gaze Interaction



Figure 4.4: Second Life’s semi-transparent control panels can be moved and resized to suit the user. The left-most panel performs camera control functions and the right-most performs locomotion controls. They are used via the mouse with clicks and drags.

These control requirements can be summarised by control source and task category, see Table 4.1, and allows us to determine what control combinations are required to interact within Second Life.

<i>Task category</i>	Control source		
	<i>Mouse</i>	<i>Keyboard</i>	<i>Speech</i>
<i>Locomotion and camera movement</i>	✓	✓	✗
<i>Object manipulation</i>	✓	✗	✗
<i>Application control</i>	✓	Partial	✗
<i>Communication</i>	✗	✓	✓

Table 4.1: Control requirements for task categories

4.3 Pilot Study 1 - Initial Assessment of Gaze Interaction

It is important to make an initial assessment of gaze based mouse emulation as a satisfactory solution for disabled users interacting with MUVEs and MMORPGs. The assessment would determine if this alternative means of interaction would enable successful, rapid and efficient interaction by giving users full control over their avatar.

4.3.1 Participants and Apparatus

In this initial assessment two expert users of eye tracking and Second Life were chosen to attempt gaze control with the online environment. The participants sat approximately 60cm from a 17" monitor with an SMI REDII remote infra-red eye tracker positioned below the screen. The mouse cursor was updated by the gaze position and dwell clicker software¹ was used to generate the mouse events.

4.3.2 Procedure

As a baseline, the participants also interacted with the environment using a normal desktop mouse. Five tasks were constructed from the task categories (see, Table 4.1), of which the avatar was required to perform a short set of actions. The tasks were as follows:

- Locomotion - The participant was required to walk the avatar along a path approximately 2 paces wide around a park, negotiating past trees and other distracting obstacles.
- Camera Movement - The participant was to move the camera from behind the avatar to face it, and then move overhead to view the avatar from above.
- Object manipulation - The participant was to create a cube and resize to be as close to 2m cube as possible.
- Application control - The participant was to change the appearance of the avatar by making the hair colour blonde.
- Communication - The participant was to chat with another avatar, generating the following sentence: 'The weather here is nice, it is always sunny and warm'.

Completion times together with the errors occurring during the task were recorded, and the subjects were asked to make comments on gaze controlled task areas they found easy or difficult.

4.3 Pilot Study 1 - Initial Assessment of Gaze Interaction

	Control source; task time (s); error count	
<i>Task domain</i>	<i>Mouse (baseline)</i>	<i>Gaze</i>
<i>Locomotion</i>	48s (3 errors)	88s (4 errors)
<i>Camera movement</i>	50s (3 errors)	122s (10 errors)
<i>Object manipulation</i>	35s	71s (3 errors)
<i>Application control</i>	20s	194s (4 errors)
<i>Communication</i>	60s (11 wpm)	224s (8 errors, 3 wpm)

Table 4.2: Pilot Study 1: Task times and error counts based on task domain

4.3.3 Analysis

The mean results can be seen in Table 4.2.

Four main types of difficulties were identified and subsequently used to define errors and were defined as follows:

- Path deviation - This occurs when the avatar has left the path that was prescribed which, contained a number of right angled turns. In the gaze condition, the steering of the avatars motion became very sensitive and more difficult to control. However this was also true in the mouse condition. One deviation from the path was counted as one error.
- Distraction - This is typically caused by the user looking at an object (being distracted) when gaze was being actively used to do something else. This would cause erratic control on the object. This error occurred typically during the locomotion task. When the eyes are being used for continuous fixation on a few locations for steering the locomotion path (even in the region of 5 - 10 seconds), it is difficult not to become distracted. For instance, it is hard not to look at a suddenly appearing object which, you can see in the peripheral vision. Glancing at the distracting object (which may be an unconscious reaction) results in changing the avatars path of motion towards the object. Every occasion of this was scored as one error.
- Selection/Accuracy - This is defined as the activation of a control object other than the one intended. Many of these were due to the fact that the controls

¹Dwell Clicker by Sensory Software: www.sensorysoftware.com/dwellclicker.html

4.3 Pilot Study 1 - Initial Assessment of Gaze Interaction

designed for mouse were too small for eye pointing; when trying to select a control an adjacent one was selected in error. Each occasion of erroneous selection was counted as one error.

- Feedback - This occurs when the eyes are looking at a control object to activate, but the effect of that control action appears somewhere on the display. The user has to look at the feedback to know whether the control action has been successful. Then the eyes move rapidly between the control point and the point of feedback. This type of error was common during the camera control task. The camera control panel was activated with a button down event but the user would immediately look at the avatar to see what the resultant effect of the camera move was; as the button was still down this caused unwanted effects on the camera position.

The following discusses the results and examine the effectiveness of gaze control against the baseline of the hand mouse for each of the task categories.

Locomotion

The major issue was of distraction, where the gaze of the user was pulled away, even temporarily, from the desired destination to some other in-world object, hence moving the avatar toward that object involuntarily. This is illustrated in a sequence (Figure 4.5, left to right) where the gaze (shown as a red star) is distracted by a tree, resulting in the avatar walking into the tree instead of staying on the path.



Figure 4.5: Pilot Study 1: Locomotion and distraction

4.3 Pilot Study 1 - Initial Assessment of Gaze Interaction

Examining the comments made by the evaluators, gaze driven locomotion was regarded as viable: ‘Steering by eye worked well, all I needed to do was to look at the object I wished to walk towards’, ‘I feel that this could be a very rapid means of steering, just look and you go right there, it could be better than a mouse’. However, the main issue of distraction was noted ‘You can’t look around while you are walking without walking off the path’.

Camera Movement

There were a large number of feedback errors caused by the difference in screen location between the camera control and the view of the avatar on the screen. This is illustrated in a sequence (Figure 4.6, left to right) with continual gazes back and forth between the camera control and the avatar (to determine the new camera position) resulting in the camera tracking and responding to the movement of the cursor - note how the camera orientation rapidly moves between the user looking at the control (left and right images), and the user looking at the avatar (centre image).



Figure 4.6: Pilot Study 1: Camera movement and feedback

Object Manipulation

Of particular interest was when once the handles of the object were acquired, the object could be resized very effectively with gaze due to the appearance of rulers extending from the object, with the user simply needing to gaze at the required measurement (placing the cursor on that measurement) for the object to be correctly resized. This is shown where the user is gazing at 4m on the ruler thus resizing the object to 4m (Figure 4.7).

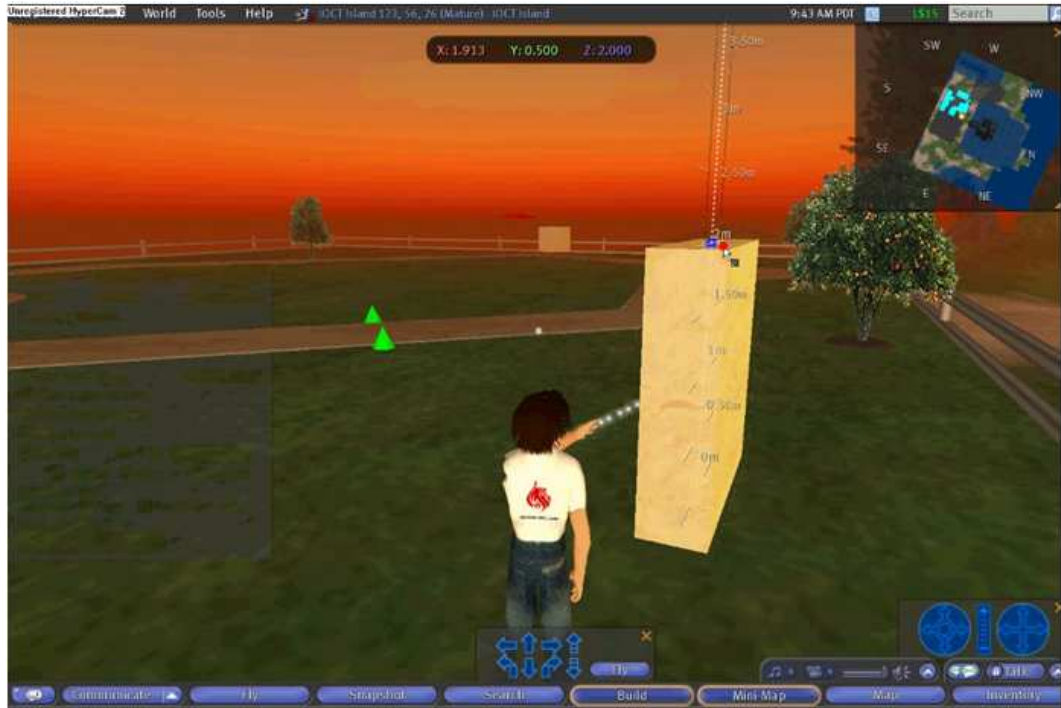


Figure 4.7: Pilot Study 1: Object sizing by gaze

Communication

The use of an on-screen virtual keyboard was slow for the mouse (11 words per minute) and very slow for gaze (3 words per minute), thus presenting a significant communication problem for our disabled user. This is illustrated by red lines (gaze paths) and red dots (gaze fixations) showing the many gazes between keyboard and chat box on Second Life when writing only 10 characters (Figure 4.8).

4.4 Discussion and Conclusions

The ideas, concepts and data discussed in this Chapter form the basis for preliminary observations. It is shown from the results and difficulties found, that using eye-gaze is constrained by the existing interfaces and will not deliver the bandwidth of interaction necessary to allow disabled users to fully participate in MUVE's and MMORPG's. At present existing virtual world interfaces will almost always force the disabled user to not perform as well as more able users, and hence fail the avatar Turing test. The types



Figure 4.8: Pilot Study 1: Feedback while eye typing

of difficulties found suggest a need for a lightweight mechanism whereby gaze control can be activated and deactivated quickly and effortlessly, with direct gaze manipulation of locomotion and objects within the virtual world.

The following two chapters take a bottom-up and top-down approach that attempt to first benchmark gaze interaction techniques and also define the game task requirements through task analysis.

Chapter deliverables:

- *An extension to Hand' taxonomy of main control and manipulation areas found in virtual environments.*
- *A pilot study that indicated gaze based mouse emulation is possible but at great cost in terms of task time and number of errors made.*

Chapter 5

Bottom Up Approach: Gaze Interaction Techniques

The aim of this Chapter is to take a bottom-up approach by using iterative design and development as a way to implement gaze interaction techniques. This process provides information by benchmarking gaze-based interaction techniques against those using keyboard and mouse in addition to providing basic usability information. Further, it attempts to define interaction technique requirements and how these can be used to develop interaction technique variants.

This Chapter is based on the following conference publications. Some portions are reproduced as are but much is rewritten to suit the flow of the Chapter and the thesis. All authors have made essential contributions and a brief note on my personal contribution is shown here.

Istance, H.; Bates, R.; Hyrskykari, A. & Vickers, S. **Snap Clutch, a Moded Approach to Solving the Midas Touch Problem.** *Proceedings of the 2008 Symposium on Eye Tracking Research and Applications; ETRA 2008*, 2008

I wrote parts of the literature review and discussion.

Vickers, S.; Istance, H.; Hyrskykari, A.; Ali, N. & Bates, R. **Keeping an Eye on the Game: Eye Gaze Interaction with Massively Mul-**

tiplayer Online Games and Virtual Communities for Motor Impaired Users. *Proceedings of the 7th International Conference on Disability, Virtual Reality and Associated Technologies; ICDVRAT 2008.* 2008

The experiment was coordinated by Howell Istance and Aulikki Hyrskykari at the University of Tampere in Finland. The software was written by myself and I wrote the full version of the paper equally with Howell Istance.

Istance, H.; Hyrskykari, A.; Vickers, S. & Chaves, T. **For Your Eyes Only: Controlling 3D Online Games by Eye-Gaze.** *Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: INTERACT 2009.* 2009

The experiment was coordinated by Howell Istance and Aulikki Hyrskykari at the University of Tampere in Finland. The software was written by myself and I wrote parts of the literature review and discussion.

Vickers, S.; Istance, H. & Hyrskykari, A. **Selecting Commands in 3D Game Environments by Gaze Gestures.** *5th Conference on Communication by Gaze Interaction; COGAIN.* 2009

The software was written by myself and the paper was written equally by all authors.

Istance, H.; Vickers, S. & Hyrskykari, A. **Gaze-based Interaction with Massively Multiplayer On-line Games.** *CHI EA '09: Proceedings of the 27th International Conference Extended Abstracts on Human Factors in Computing Systems.* 2009

The paper was written equally by all authors.

5.1 A Software Device for Interpreting Gaze Data

There are two approaches that can be used to allow eye gaze interaction with a game. Firstly, the game can be modified or designed to recognise gaze input directly and secondly, eye gaze can be used as a mouse or keyboard emulator. The approach used

5.1 A Software Device for Interpreting Gaze Data

here is based on mouse and keyboard emulation with the game client reacting as if the events came directly from a mouse or keyboard. The intention is that the software is not dependent on a particular game and so sits independently of the game client and is easily configurable to emulate different sequences of mouse and keyboard events. In this way it can be used with almost any game that operates using mouse and keyboard. However, developing middle-ware such as this is not without its difficulties.

The first major obstacle to overcome is the game architecture that the middle-ware is interfacing with. Emulation based middle-ware works by sending key and mouse events into the Windows message queue. This works well if the target game takes input messages from the same queue but some games do not, such as those that make use of Microsoft's DirectInput. Rather than taking input messages from the Windows queue, games that use DirectInput take messages direct from input devices. To overcome this it would be necessary to either modify the game directly or use a virtual device driver. Similar to virtual CD/DVD device drivers, it would be possible to route gaze input through a virtual keyboard/mouse driver which generates input events. DirectInput will then recognise these as true events as if from a mouse or keyboard.

The second major obstacle is software firewalls. Many online game developers utilise some form of firewall to prevent games being controlled by bots. For instance, the MMORPG Rappelz¹ which, is similar to many others in that player's must invest a large amount of time in developing and advancing their character through completing quests and killing monsters. Some players who do not wish to invest this time will attempt to control their character automatically by creating a bot. The bot is typically controlled by a piece of middle-ware that may use pixel recognition to determine what is happening on the player's screen and then automatically move the character to continuously search for, kill, steal gold from monsters. Player's can leave their computers on for days at a time and when they return their character will have advanced in level and amounted a large sum of in-game currency. To protect from this form of malicious game-play, Rappelz incorporate a software firewall called GameGuard² that blocks input events which have not been generated directly from an input device.

Due to these two potential difficulties, gaze based middle-ware is restricted to games that do not use DirectInput or incorporate a software firewall. Despite these two

¹Rappelz MMORPG: <http://rappelz.gpotato.com/>

²GameGuard: <http://gameguard.nprotect.com/>

limitations there remain many games that can be interacted with.

5.1.1 Modes of Interaction

Most gaze control systems have a facility to pause or suspend active control. However, this is often not used as the cognitive overhead in doing so is prohibitive. One element in a solution therefore is to incorporate a very lightweight process to turn gaze control off. Another is to support several modes of mouse, keyboard or joystick action simultaneously and to have a similarly light weight way of switching between these. One such method is by using glances and detecting when a glance or ‘flick’ of the eye off the edge of the screen occurs. Each directional glance could represent a different mode of operation. The use of modes in the user interface is not a new concept and was once considered bad practice due to the overhead of remembering which mode was currently in operation (Nievergelt and Weydert, 1987). There are also potential difficulties in the user remembering where each mode is located and how to switch between them. Therefore, for moded operation to be successful, there are a number of design requirements that should be met such as:

- Movement and actions needed to accomplish the task should be efficient.
- Low effort associated with using the mode and with changing between the modes.
- Easy to remember how the modes work and how to activate them.
- Clear feedback about the mode the user is currently in.

Here, different interaction techniques are considered to be different modes of interaction. A traditional gaze based interaction techniques such as ‘left mouse dwell click’ in which, the cursor follows the user’s gaze position on screen and when the user dwells at a point on screen long enough a left mouse click event is sent to that point, can be considered one such mode.

5.1.2 Architecture and Implementation

The gaze based middle-ware developed here has four modes available to a user at any one time; a mode collection. In order to change between each mode the user need simply glance off one of the four edges of the screen. The user is notified by a mode

5.1 A Software Device for Interpreting Gaze Data

change with speech (if activated) and a green bar that sits on the screen edge that they have just looked off. Additional feedback is given by changing the system cursor but this is unreliable as some games will define their own cursors. Usually one of the four active modes is ‘off’ where dwelling on the screen has no effect.

The middle-ware was developed in C# and was given the name Snap Clutch, see Figure 5.1. The original concept was a method of temporarily disengaging eye control by glancing off-screen, or ‘de-clutching’ eye control, as a means to overcome the Midas touch problem (Istance et al., 2008). It was developed using the Tobii SDK but is also compatible with the Eye Tracking Universal Driver¹. The architecture of Snap Clutch

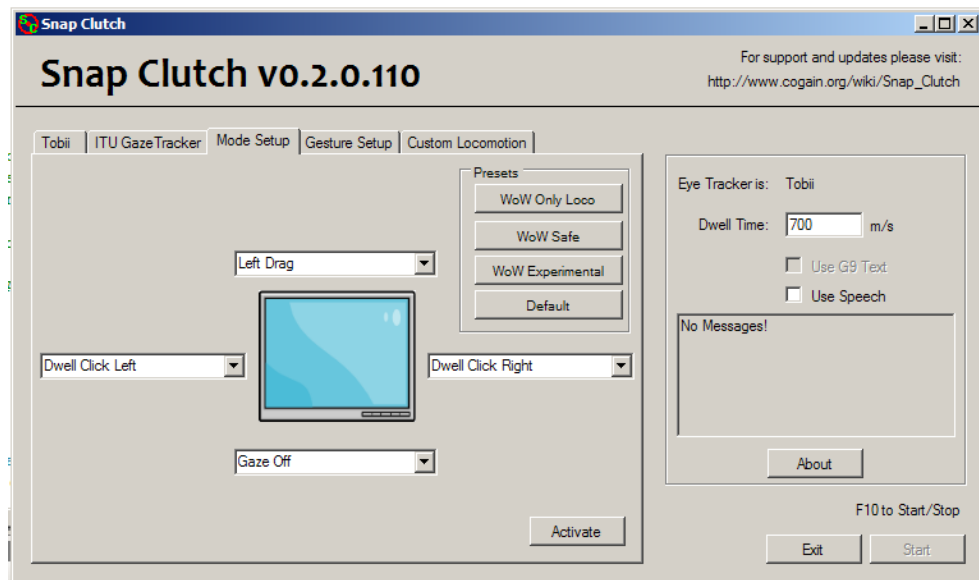


Figure 5.1: Screenshot of the Snap Clutch interface. Here, different modes are assigned to edges of the screen using drop down control.

is such that additional interaction modes can quickly be created and added for use by the application, see Figure 5.2. When inherited by the interaction mode being created, the Mode class has three virtual functions that can be overridden:

- virtual MOUSESTATE Execute(Point gazePoint, MOUSESTATE mouseButton-State) : *This function is used to define how the interaction technique uses the gaze position. For instance, in a mouse emulator this function would update the cursor position.*

¹ETU Driver: <http://www.cs.uta.fi/oleg/etud.html>

5.1 A Software Device for Interpreting Gaze Data

- virtual MOUSESTATE SomeEvent(Point gazePoint, bool switch) : *This function is used to define what happens when a click or switch event has occurred.*
- virtual MOUSESTATE BlinkEvent(Point gazePoint, int eyeRef) : *This function is used to define what happens when a left or right blink event has occurred.*

MOUSESTATE is a struct that is used to store the different states of all mouse or key events that are used. Modes are gathered into Mode Collections that are unique to a particular environment or type of game. The Snap Clutch Component Module manages the modes so that the correct events are being generated. At the core of the architecture is the Gaze Engine that processes all of the raw gaze data from the eye tracker. This class filters noise and performs some smoothing of the current gaze position. Several events are registered, including fixations, on and off screen dwell, off screen glances and left/right winks/blinks.

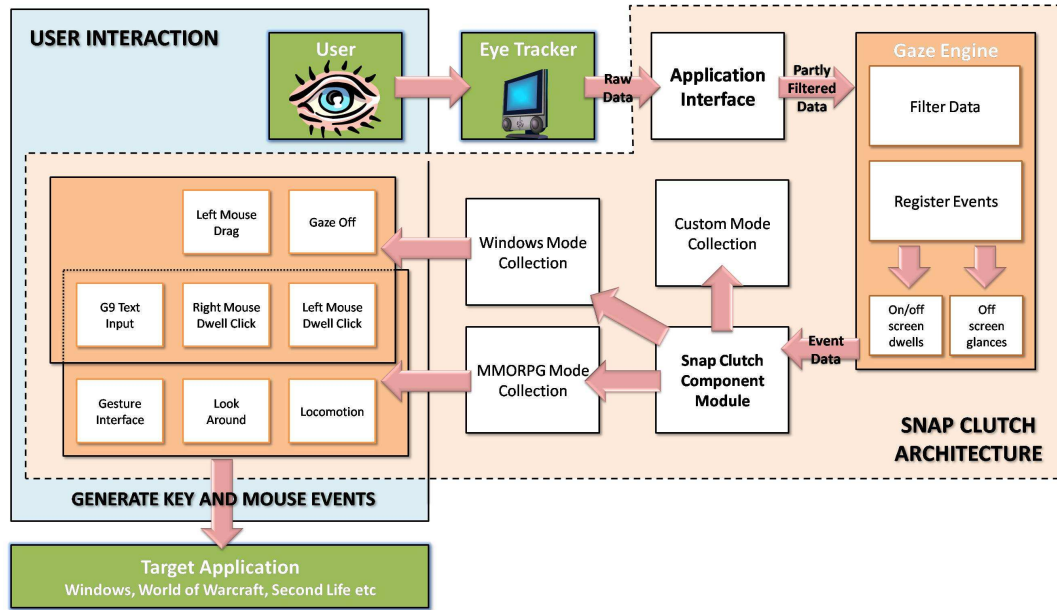


Figure 5.2: Snap Clutch flow of operation.

Gaze Engine Definitions

The definitions of smoothing and fixations implemented within the Gaze Engine are as follows. They are based on an eye tracker working at 60HZ.

5.2 Task Categories and Novel Gaze Interaction Techniques

Smoothing

A rolling average was used to determine smoothed gaze position. The buffer of the rolling average array is 30 and so there is initial latency of 480ms. Unfiltered (X, Y) gaze data points are entered into the array and after the initial latency, the average value of the last 30 data points is returned. Thus, the estimated gaze position will be the average of the last 480ms of gaze data points.

Fixation

Fixations are periods where the gaze position is relatively stable. They are calculated using raw gaze data points and not smoothed data points. A fixation is determined as taking place when there are five consecutive (X, Y) gaze data points within a region of 30 pixels of the previous data point. Thus making a fixation of at least 80ms in length. There is no mechanism implemented for detecting fixation drifts that will occur.

5.2 Task Categories and Novel Gaze Interaction Techniques

The broad task categories discussed in Section 4.2 are:

- Locomotion and camera movement
- Object Manipulation
- Application Control
- Communication

The pilot study demonstrated that performing these tasks using gaze based mouse emulation was not adequate due to the inability to deactivate gaze control and also the size of the interface controls. There is also a need to maintain visual attention to parts of the screen even when the interaction controls exist elsewhere. Finally, there is a requirement to perform all of the tasks in real-time as required by the environment.

The following Sections take these task categories and discuss possible gaze interaction techniques that may allow them to be performed.

5.2.1 Locomotion

On-screen Dwell Keyboard

Locomotion in environments such as Second Life can usually take place using the arrow cursor keys and/or the ‘W, A, S, D’ cluster of keys. Perhaps the most obvious way in using gaze to perform this task is to use a dwell activated on-screen keyboard that display only these keys. However, this solution has several problems, firstly, the keyboard will obscure some of the viewing area on screen, and secondly, it is necessary to maintain gaze consistently on the buttons. Finally, there will be latency with each change in direction made that is equal in length to the dwell time.

Drag-from-here

This interaction technique makes use of a transparent control panel similar to that found in Second Life (see figure 4.4). The technique allows the user to leave the cursor at a position (by a dwell), then start a mouse drag action from that position by moving the gaze point. The drag action can be finished by performing a second dwell or by changing the mode. Feedback of this technique before the drag action is by a green up arrow and during a drag action changes to a red up arrow, see Figure 5.3. The initial dwell to drop the cursor must be on the forward control of the transparent control panel. The first problem with this is that the initial dwell must be accurate enough to be placed in the correct place upon the control panel. Secondly, the avatar will walk towards whatever is being looked at. Finally, if an object in the distance is gazed upon for too long then a dwell would occur releasing the drag function.

Transparent Overlay

An alternative to using a specific interaction technique is to have a dedicated task ‘mode’ that can be used just for that one task, e.g. a locomotion mode. As soon as the mode is activated the character starts moving forward, then in order to change direction they look to the left and right of the screen. Transparent overlays are on-screen gaze sensitive regions that are not visible to the user. When the user’s gaze enters one of the regions then a series of keyboard or mouse events are sent to the game, see Figures 5.4 and 5.5. An advantage of this technique is that a user does not have to steer a cursor or have a high degree of eye tracking skill. They need only look around the 3D world

5.2 Task Categories and Novel Gaze Interaction Techniques

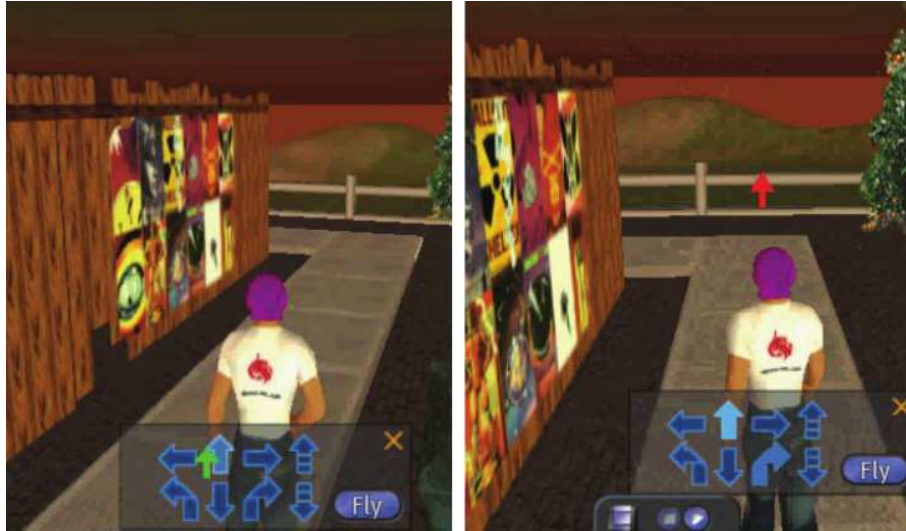


Figure 5.3: ‘Drag-from-here’ cursor being used for locomotion. The interaction technique is in it’s inactive state (left) and active state (right) (Istance et al., 2008)



Figure 5.4: Transparent Overlay: ‘Look around’ task mode. The regions are not visible to the user. In this mode, the character cannot walk but will rotate left or right whenever the region is activated. The amount of rotation is proportional to the distance from the centre.

as they normally would with the exception that their character moves based on their current gaze position. The transparent overlays used in these two modes are activated when a single fixation is detected within one of the regions. The positioning of these zones in relation to movement was derived from the pilot study in Section 4.3. Having

5.2 Task Categories and Novel Gaze Interaction Techniques



Figure 5.5: Transparent Overlay: ‘Locomotion’ task mode. The regions are not visible to the user but are activated when their gaze fixates within them. In this mode, the character can walk, rotate left and right and walk backwards whenever the associated region is activated.

the avatar walk towards the direction of the cursor (or gaze position) was immersive but would result in the character always walking towards everything that they looked at. Therefore, these zones were positioned so as to give a safe area for the user to look at without the character always walking towards their gaze position.

5.2.2 Camera Movement

Park-it-here

As with the ‘drag-from-here’ interaction technique, the ‘park-it-here’ technique makes use of a control panel. When active, dwelling on a position allows the user to leave the mouse pointer at that position. The user is free to look elsewhere and a dwell action causes a mouse down and up event at the pointer position. The user can then pick the cursor up again just by looking back at it. It can be dropped at some other point by dwelling at that point. The cursor is green when being moved by the gaze point, and red when parked and active, see Figure 5.6. This technique eliminates the feedback problems encountered during the pilot study, however it does require accurate placement of the cursor on to the panel control.



Figure 5.6: ‘Park-it-here’ cursor being used for camera control. The interaction technique is in it’s inactive state (left) and active state (right) (Istance et al., 2008).

5.2.3 Object Manipulation and Application Control

The main problem encountered for these two task categories during the pilot study was due to accuracy of the eye tracker and the small size of the interface controls. Therefore, they are discussed together.

Magnifying Glass and Toolglass

The magnifier enlarges the associated portion of the screen so as to allow for a more accurate placement of the toolglass. Once the toolglass has been dropped (by dwell), a red cursor is placed at the user’s gaze position. The user can then choose to send either a single left click, double left click or right click to the interface control beneath the magnifier. The choice is initiated by the user looking off either the top, left, or right of the toolglass. See Figure 5.7.

A simpler version of the magnifier toolglass is the magnifying glass, see Figure 5.8. Here, as before, the magnifier enlarges portion of the screen and is dropped by a dwell. Once dropped, a red cursor is placed at the user’s gaze position and the user can then dwell on the enlarged interface components in order to select them. A significant problem with this technique is the number of additional steps required in order to select a control.

Gestures

A similar gesture implementation to Quikwriting is used with the gesture zones located around the avatar. The zones are visible, but nearly transparent. There are only 4



Figure 5.7: Magnifying Toolglass. The user drops the toolglass over a magnified control using dwell; then looking off the left of the toolglass generates a left mouse click, the top generates a double left mouse click, and the right generates a right mouse click.



Figure 5.8: Magnifying Glass (Istance et al., 2009). The user drops the toolglass over a magnified control using dwell; the user then controls a red cursor and is free to dwell on the enlarged interface controls.

active zones allowing a total of 12 recognisable gestures; 4 x two-stroke gestures and 8 x three-stroke gestures. The size of the overlay area containing the active zones is configurable to accommodate individual preferences for camera angles and fields of view. Key and mouse events are allocated to each of the gestures. The appearance of the World of Warcraft game client window with the active areas visible is shown in

Figure 5.9.



Figure 5.9: Performing a gesture in World of Warcraft, to select Spell A in this case

There is no labelling on the zones themselves, but a small reminder overlay showing which commands are currently associated with the 12 gestures can be made visible and located over any part of the client window (visible on the right in Figure 5.9). The figure shows a three-stroke gesture that selects the upper left element associated with the bottom left zone (zone 4A, command ‘spell A’ see, figure 5.10), in a similar way to the Quikwriting scheme.

A significant problem with this gesture implementation is that it is necessary to learn 12 different gestures and their associated functions. Further, 12 gestures may not be enough for the full range of interface controls that may be necessary. It does however, allow the user to keep their gaze focused on the centre of the screen where much of the attention would lay. Additionally, the gesture zones are quite large and so it is not likely that accuracy problems will be encountered.

5.2.4 Communication

G9 Predictive Text

The pilot study demonstrated that communication using an on-screen keyboard resulted in a low word count; much of the screen being obscured by a large, full keyboard; and

5.2 Task Categories and Novel Gaze Interaction Techniques

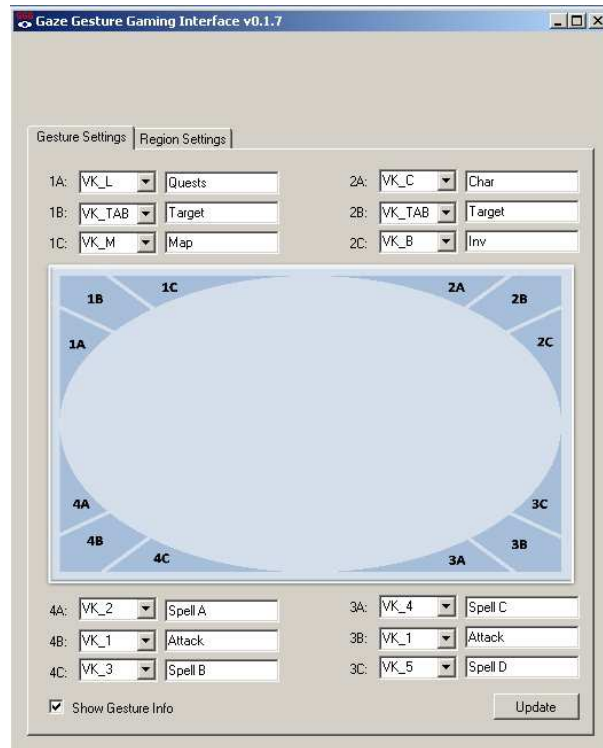


Figure 5.10: Gesture configuration dialog, showing the command generated in Figure 5.9

poor text feedback to the user. Therefore, a smaller semi-transparent keypad that looks and works in a similar way to T9 texting on a mobile phone was implemented, see Figure 5.11. As any communication that takes place in an online world is likely to be based on speed and minimal text (so, similar to texting, lots of abbreviations and lower case letters), it is not necessary to provide a full set of characters. The most likely word or part-word that matches the string the user has entered is displayed above and below the centre of the key that currently contains the gaze point. This is highlighted as soon as the gaze point enters it. The ‘delete’ key (top left) removes characters from the string. Other matching words or part words from the dictionary are cycled through by looking at the ‘cycle’ key. The ‘back’ key sends backspace characters to the buffer and the ‘start/finish/enter’ key sends a return character to the buffer. This sends the entire string to the game. Only one punctuation mark (‘?’) is supported and there are no upper case characters.



Figure 5.11: G9 Predictive Text. No cursor/feedback of gaze position is given to the user however, the buttons are dwell activated so when their gaze enters one of the button regions then a visible dwell pulsing cursor timer is shown.

5.3 Modifying Gaze Interaction Techniques

The interaction techniques presented are implemented as described; that is, single and fixed modes. However, there is a requirement for modes to implement different variations of the same technique. One example of this is the ‘locomotion’ and ‘look around’ mode in which they are both variants of the transparent overlay interaction technique. This Section attempts to classify the possible variants based on the interaction technique requirements of the user and the hardware.

Each technique can be modified in accordance with the following taxonomy:

- **Complexity:** There are two types of complexity: firstly, there is complexity related to fulfilling the interaction technique requirements, e.g. number of gesture steps; secondly, there is complexity related to the number of available choices, e.g. number of possible gestures.
- **Accuracy Requirement:** The size of the interaction area, measured in degrees as

5.3 Modifying Gaze Interaction Techniques

the visual angle and assuming a viewing distance (D) of 600mm, see equation 2.1; e.g. increasing/decreasing the size of an on-screen button.

- Spatial position: The position of the interaction area; e.g. positioning all buttons to one area of the screen such as to the left
- Temporal: The length of time required to initiate one step of the interaction technique; e.g. a dwell count
- Intelligent: Calculating calibration offsets, part automation of interaction techniques.

By applying these modifying methods it is possible to expand our range of implemented interaction into many possible derivatives such as those found in Table 5.1¹. A visual representation of this can also be seen in Figures 5.12, 5.13, 5.14 and 5.15. Each technique is shown grouped by it's relative parent technique type.

Taking *Transparent Overlay* as an example: there are many possible variants with two possibles shown for this technique. They can both be operated using a single gaze fixation; a simple variant may have 3 available functions and the other 6; the smallest zone for the simple variant may require 10.5° of accuracy and the other 7.5°; commands can be initiated after 80ms of interaction. Visual representation of how this technique may look can be seen in Figure 5.16 which, also demonstrates the ability of spatial positioning with the overlay components on the simple variant being positioned differently.

In addition to considering complexity, accuracy, spatial, temporal and intelligent as properties for modification, they can also be considered user and hardware requirement properties. For instance, by using the *Mouse Emulation* interaction technique we can assume that it is being used for 2D desktop interaction. Thus, a need to interact with small interface components, so from the Table a user and their hardware must be capable of achieving 1.3° accuracy; the complete desktop interface are possible options for interacting with so the user must be cognitively aware of exactly what they are intending to do; only two choices are available for the actual interaction technique- 'move cursor' and 'select'; and finally, the user is required to perform at least 800ms of 'interaction' in order to fully use the technique (a dwell click).

¹Note: Spatial position has been omitted as this modification relates to many possible positions that each interaction area can be used.

<i>Parent Type</i>	<i>Interaction Technique</i>	<i>Complexity Steps</i>	<i>Complexity Choices</i>	<i>Smallest Zone</i>	<i>Accuracy</i>	<i>Temporal</i>
<i>Transparent Overlay</i>	Overlay (simple)	1	3	420x1050px	10.5°	80ms
	Overlay	1	6	400x300px	7.5°	80ms
<i>Gesture</i>	2 Step	3	8	700x700px	17.5°	400ms
	2 Step Delayed	3	8	700x700px	17.5°	1000ms
	3 Step	4	16	700x700px	17.5°	750ms
<i>Dwell</i>	Button Large	1	4	400x400px	10°	800ms
	Button Large Delayed	1	4	400x400px	10°	1200ms
	Button Medium	1	15	200x200px	5°	800ms
	Button Smallest	1	30	100x100px	2.5°	800ms
<i>Cursor Based</i>	Toolglass	2	15..n	300x300px	7.5°	1200ms
	Mouse Emulation	2	10..n	50x50px	1.3°	800ms
	Zoom Select	2	10..n	300x300px	7.5°	800ms

Table 5.1: Example of how the gaze interaction techniques may be modified.

5.3 Modifying Gaze Interaction Techniques

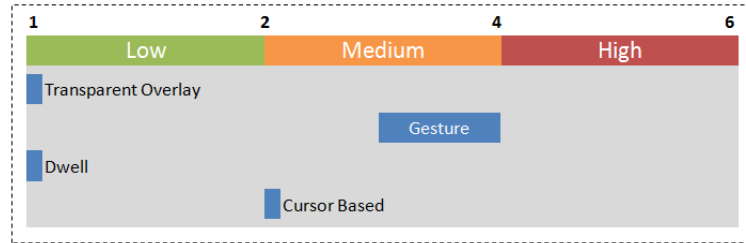


Figure 5.12: Complexity Choices: The scale shows the range of number of steps required to perform the technique.

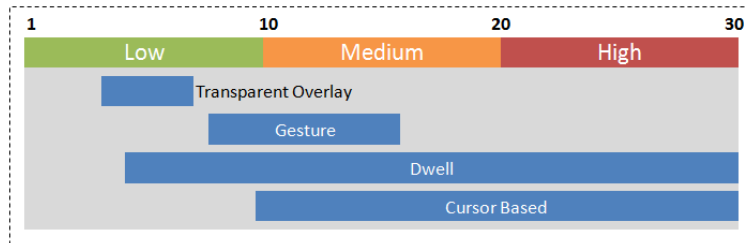


Figure 5.13: Complexity Functions: The scale shows the range of number of functions/choices available for use.

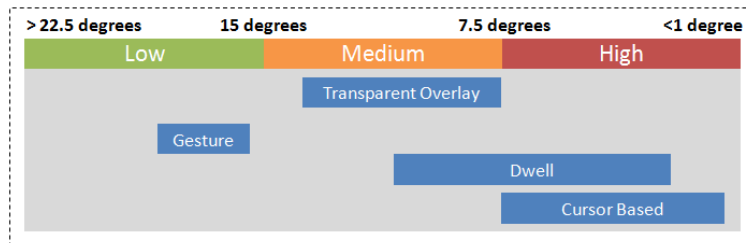


Figure 5.14: Accuracy Requirements: The scale shows the range in size of the interaction area.

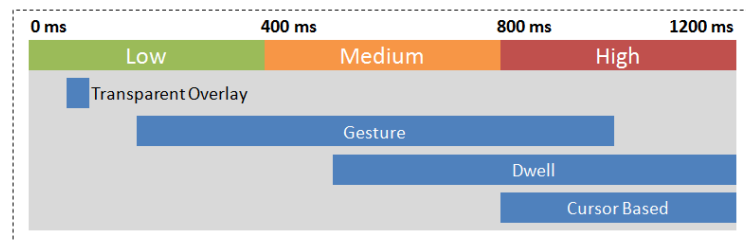


Figure 5.15: Temporal Requirements: Classifying the parent interaction techniques in terms of it's temporal demands.

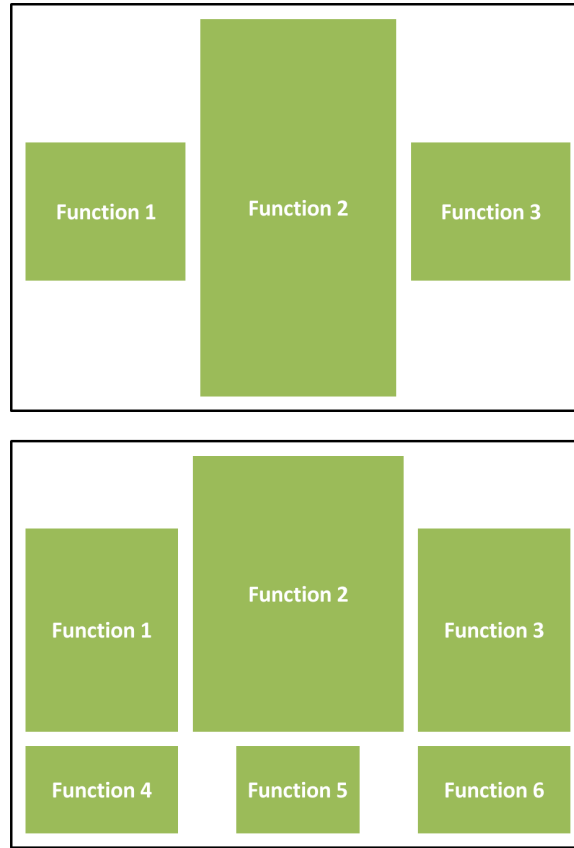


Figure 5.16: Two variations of the transparent overlay interaction technique from Table 5.1: Overlay (simple) (top) and Overlay (bottom)

5.4 Evaluation Study 1: Gaze Interaction in Second Life

The first pilot study (see Section 4.3) demonstrated that using Second Life with gaze-based interaction techniques resulted in distinctively longer task completion times than when using conventional interaction techniques. In order to achieve parity of gaze interaction with normal keyboard and mouse, it is necessary to be able to identify the usability issues with gaze control. This will generate understanding of what influences the speed of interaction (time of task completion) and the type of errors that are made.

The partitioning of task time into productive time and error time has long been a feature of usability engineering (Gilb, 1984). The time spent in a specific error condition represents the potential time to be saved in task completion if the cause of that error can be removed. This relative saving in task completion times by addressing each type

of error represents a kind of cost saving benefit of redesigning different features of the user interface.

5.4.1 Participants and Apparatus

The study involved twelve participants aged from 20 to 56 (mean = 29). Ten of them were students and two were university lecturers all from the University of Tampere in Finland, who were experienced users of gaze interaction. All participants were able bodied. The study was carried out using a Tobii T60 screen integrated eye tracker and all tasks captured by video. The modes used by Snap Clutch were:

- Glance up - Look around, see Figure 5.4
- Glance left - Left button click (a gaze dwell causes a left button click)
- Glance right - Right button click (a gaze dwell causes a right button click)
- Glance down - Locomotion, see Figure 5.5

The ‘look around’ and ‘locomotion’ modes used were early implementations of the transparent overlay technique in which there was no control in rotational speed.

5.4.2 Procedure

Two sets of three tasks were devised and carried out within a purpose built Second Life environment that represented the Gateway House at De Montfort University. They were as follows:

1. Locomotion: The participant was required to walk from the main entrance, up the staircase and go into a room where there were display panels about individual university modules, see Figure 5.17. The task required the participant to retrieve a module code from a particular panel. Each participant set were required to retrieve a different module code from a different panel.
2. Object manipulation: The participant was required to change a slide or a web page from within the main lecture theatre and accept the change using a dialog box. One subject set changed the presentation slide by ‘touching’ a button on a control panel located on the stage. The other participant set caused a web page

5.4 Evaluation Study 1: Gaze Interaction in Second Life

to be displayed by ‘touching’ another button that was located near the stage. To achieve this task both sets of subjects were required to perform a right click to display the Second Life pie menu (see Figure 4.1) and then a left click to select the ‘Touch’ menu item.

3. Application control: The participant was required to change the appearance of their avatar. One participant set was to remove the moustache and the other was to raise the height of the eyebrows. To achieve this, the participant was required to use the Second Life pie menu but with the selection of the ‘Appearance’ menu item. This caused a dialog box to appear and the subject then had to select a group of features to edit from a panel of vertical buttons. A horizontal slider was used to change the selected feature.



Figure 5.17: Evaluation Study 1: Screenshot of a participant performing the locomotion task in Second Life.

The twelve participants were split into two groups of six. One half did one task set with the keyboard and mouse, followed by the other task set using gaze control. The other half started with gaze control followed by keyboard and mouse. All subjects with the exception of the two experienced participants had not used Second Life before, nor did they have any previous experience of using gaze control. Each participant was given a fifteen minute introduction to Second Life followed by a fifteen minute introduction to using gaze control and Snap Clutch. This involved a series of simple training exercises.

Each task set began with the avatar in the same place, standing at the entrance to the building. Tasks were completed in the same order for all subjects: locomotion; object manipulation; application control. The task was first explained and then the participant was asked to complete it. If required, the participant was reminded of the task during completion. All participants were asked to complete a brief questionnaire upon completion of the two task sets. They were advised that they could withdraw from the trial at any time. Each session took between 45-55 minutes to complete.

5.4.3 Analysis and Results

Four different categories of errors were identified that the subjects made during the tasks. These were annotated upon the videos using the open source video annotation application called Elan. The video data from one subject was marked up by two people so that consistency of the outcomes could be checked. As a result there were minor adjustments made to the error categories and definitions but for the most part there was a high degree of agreement between the two analyses. The four main error categories were as follows:

1. Locomotion error: being one of the following,
 - Unintentional motion backwards (the gaze point first moves through the ‘move backwards’ zone of the screen after glancing down to change into the locomotion mode).
 - Unintentional rotation left or right (the subject means to glance off screen to change modes when in ‘no control’ mode, but rotates the avatar instead).
 - Turn overshoot (the subject deliberately turns while in ‘locomotion’ mode but turns too far and has to correct).
 - Walk overshoot (the subject tries to stop, but the change to ‘no control’ mode takes too long and the avatar walks to far and subsequently has to reverse).
2. Mode change error: an unintentional change of the mode; a subject tries to rotate left or right, in ‘no control’ or ‘locomotion’ mode but changes mode by mistake by looking too far off screen.

5.4 Evaluation Study 1: Gaze Interaction in Second Life

3. Accuracy error: a subject tries to select a target but misses due to inaccurate pointing.
4. Misunderstanding error: a subject misunderstands / mishears / forgets what to do.

For each subject and for each task the total time spent in each error condition was subtracted from the total time, leaving us the non-error time for each trial. The outcome of the trials for the three tasks is as shown in Figure 5.17 and Table 5.2.

<i>Task</i>	<i>Condition</i>	<i>Error free Time</i>	<i>Error Count</i>	<i>Error Time</i>	<i>Error %</i>	<i>KBM/Gaze Ratio</i>
<i>Locomotion</i>	<i>Gaze</i>	57.9 sec	4.5	21.4 sec	22.7	1:1.63
	<i>KBM</i>	47.1 sec	0.18	1.32 sec	2.54	
<i>Application</i>	<i>Gaze</i>	79.3 sec	3.1	41.8 sec	29.9	1:2.82
	<i>KBM</i>	40.2 sec	0.45	2.7 sec	5.95	
<i>Object</i>	<i>Gaze</i>	18.99 sec	2.45	21.89 sec	50	1:4.6
	<i>KBM</i>	7.84 sec	0.18	1.05 sec	6.1	

Table 5.2: Mean task times, error counts, error time and error percentage for all tasks (kbm = keyboard and mouse)

Table 5.2 shows the tasks along with the average error-free time; the total error count; the total error time; the percentage of error time. This is also represented in Figure 5.18 but with the error-time divided into their associated error categories.

One of the subjects had difficulty in calibrating the eye tracker. She was able to complete all of the tasks in the gaze condition but the number of accuracy errors was far greater than the other subjects with more than 3 standard deviations from the mean. Consequently all data from this subject was removed from the analysis.

The results show that all subjects were able to complete the three tasks using eye gaze. The non-error time enables comparison of task times if the cause of the errors can be removed through design changes. The non-error times for gaze is encouraging, especially those for the locomotion task. With only a short training session, subjects would be able to complete the locomotion task using gaze almost as quickly as they would using a keyboard, if the cause of the locomotion errors could be removed. The reasons

5.4 Evaluation Study 1: Gaze Interaction in Second Life

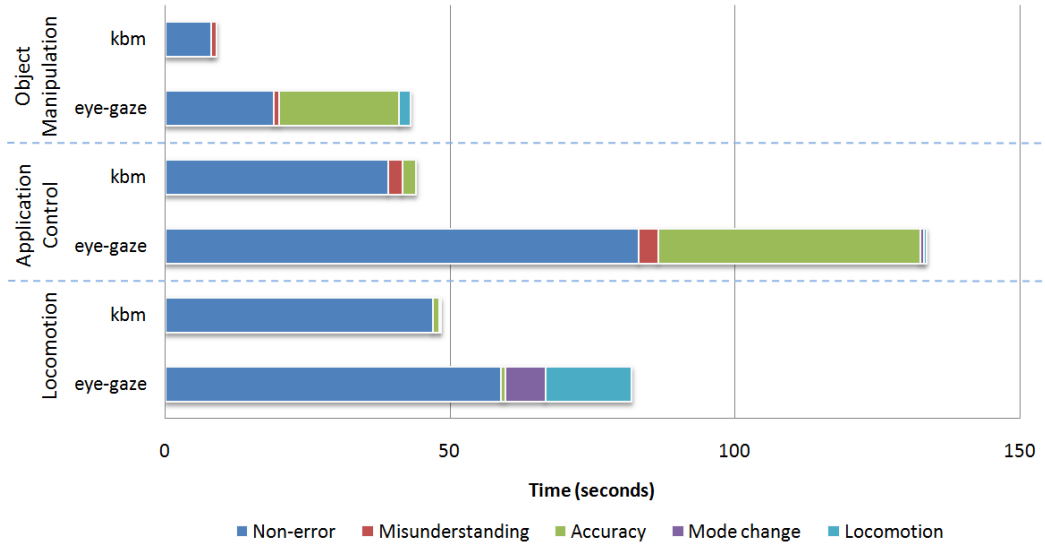


Figure 5.18: Evaluation Study 1: Task performance with non-error and error time stacked.

for the locomotion errors were partly due to the speed of movement by the avatar in response to the key commands generated by Snap Clutch. This caused overshooting and undershooting of avatar movement that would need to be corrected. The processing pipeline in using a single computer is a significant contribution to this: eye tracker - emulator (Snap Clutch) - Second Life (additionally, in the experimental condition, the video capture software). There are also possible optimisations that can be made to improve the emulator software. Another source of locomotion error is the location of the backwards motion overlay zone at the bottom of the screen, see Figure 5.5. The gaze position has to travel through this zone after changing into the locomotion mode and combined with the latency within the system an unwanted backwards movement results. These issues can be addressed within the implementation of the locomotion mode in addition to examining causes for response latency.

The biggest cause of errors in the application control and object manipulation tasks is the difficulty in hitting the small control objects within the dialog boxes. This was exacerbated by latency in generating click events probably due to the processing pipeline. One solution to reduce these types of errors is to incorporate some form of zoom interface that is common within 2D gaze driven interfaces.

5.5 Evaluation Study 2: Gaze Interaction in World of Warcraft

Further statistical analysis of the results using ANOVA is also possible, however this goes beyond the scope and questions that this evaluation study was intended to address.

5.4.4 Conclusions

The evaluation study has been successful in identifying the extent and causes of the difference in performance between the gaze and keyboard-mouse conditions. This has revealed specific design changes that address these differences and also gives an indication of the likely performance improvements. Importantly however, it has demonstrated the feasibility and potential for gaze based interaction with online virtual environments. In particular that of gaze based locomotion, in which there lies real possibility that eye gaze can achieve parity with that of mouse and keyboard when performing such tasks.

5.5 Evaluation Study 2: Gaze Interaction in World of Warcraft

The results from evaluation study 1 were encouraging and showed that all participants were able to complete all tasks, after only a brief introduction and training with Second Life and with gaze interaction. This study attempts to address the design issues and thus explore to see if non-error time performance ratios were achievable following modifications of Snap Clutch and when using World of Warcraft (as an example of a popular MMORPG).

Evaluation study 1 showed that turning using gaze was very sensitive and often caused overshooting that required a steering correction in the opposite direction. This is due to the ‘locomotion’ mode used not having any rotation control and thus addressed by varying the amount based on the ‘X’ coordinate gaze position, see Figure 5.5. A simpler version of the magnifying toolglass, see Figure 5.7, was used to counteract the accuracy problem. The user can pick up the magnifier glass by a dwell on a semi-transparent icon placed on the game window and a subsequent dwell on the magnifier icon turns the magnifier off.

Unlike the previous evaluation study in Second Life, this study included time-constrained interaction tasks with other characters. The tasks were representative of beginners’ level experience and so a character was created with a medium experience

level (level 16) and the same character was used in the same virtual space by all participants. The server used was public so there were other characters in the same space as the tasks conducted in a realistic play environment adding a reasonable amount of random distraction caused by external events in the game.

5.5.1 Participants and Apparatus

The study involved ten participants aged between 18 and 44 (mean = 25). These were 9 males and 1 female, all were able-bodied, and all were students or staff at the computer science department at the University of Tampere in Finland. None had taken part in the first experiment. Five had current extensive games playing experience with MMORPGs, three with World of Warcraft. All of the other 5 had played computer games, but did not consider themselves to be experienced MMORPG players. The study was carried out using a Tobii T60 screen integrated eye tracker and all tasks captured by video. The same mode arrangements were used in Snap Clutch as in evaluation study 1 however, their operation was modified as described above:

- Glance up - Look around, see Figure 5.4
- Glance left - Left button click (a gaze dwell causes a left button click)
- Glance right - Right button click (a gaze dwell causes a right button click)
- Glance down - Locomotion, see Figure 5.5

A fifth mode was added activated by an on-screen dwell button that would display the magnifier glass.

5.5.2 Procedure

Three tasks were designed that were representative of beginner level play in the game. These were:

1. Locomotion: The participant was required to walk to a location identified on the inset map, to turn around and return to the starting point. There was no control over character speed and the participant was asked to stay on the path and complete the task as quickly as possible.

2. Fighting: The participant was required to find and fight a level 3 monster by casting the same spell as many times as possible during the fight (by left clicking on an icon located on the shelf in the centre bottom of the screen). The difference in levels assured the participant would always win. After the fight, the participant was asked to loot the corpse (by right clicking on it) of one item (by left clicking on the list of treasure).
3. Equipment: The participant was required to put on or wear four items of equipment by opening a pouch (left clicking its icon in the bottom right of the screen); then opening the character sheet (left clicking its icon also in the bottom right of the screen); then selecting an item from the pouch (left clicking on its icon in the pouch); then selecting the highlighted slot in the character sheet which was open in the upper left part of the screen (again by left clicking in the empty slot); then closing both windows (left click in the close box in the top right of the window)

Each trial consisted of a training phase (50 to 60 minutes), a break (20 to 30 minutes), and a data collection phase (about 30 minutes). The first part of the training covered the use of the gaze device, the magnifier, and locomotion mode. The second part of the training consisted of a structured introduction to World of Warcraft and completing a set of standard tasks. After a break, all the four tasks were recorded with the keyboard and mouse. The same four tasks were then carried out using gaze. The order of conditions during the trials was not counterbalanced as it was important to increase the practice obtained before the gaze trial. There was no expectation that gaze would perform better than keyboard and mouse.

5.5.3 Analysis and Results

Table 5.3 and Figure 5.19 shows the mean completion time for each task and the keyboard/mouse to gaze ratio.

Locomotion Task

Data from one participant was omitted from the quantitative analysis but retained in the analysis of subjective data. This was due to problems calibrating the eye tracker. The keyboard/mouse to gaze performance ratio was better than expected for the locomotion task in comparison to evaluation study 2; 1:1 in comparison to 1:1.63. In

5.5 Evaluation Study 2: Gaze Interaction in World of Warcraft

Task	Time in seconds		KBM/Gaze Ratio		
	KBM	Gaze	Pilot 1	Eval. 1	Eval. 2
<i>Locomotion</i>	80.9 (2.6)	83.2 (7.5)	1:1.8	1:1.63	1:1
<i>Fighting</i>	15.1 (2.7)	32.7 (17.8)	n/a	n/a	1:2.1
<i>Equipment Pt 1</i>	3.4 (1.1)	17.2 (14.6)	1:9.7	1:2.8	1:5
<i>Equipment Pt 2</i>	17.4 (5.4)	45.4 (23)	1:9.7	1:2.8	1:2.6

Table 5.3: Pilot Study 3: Mean task completion times for evaluation study 2 with standard deviation shown in brackets; and comparison of keyboard/mouse to gaze ratio for pilot study 1 and evaluation studies 1 and 2..

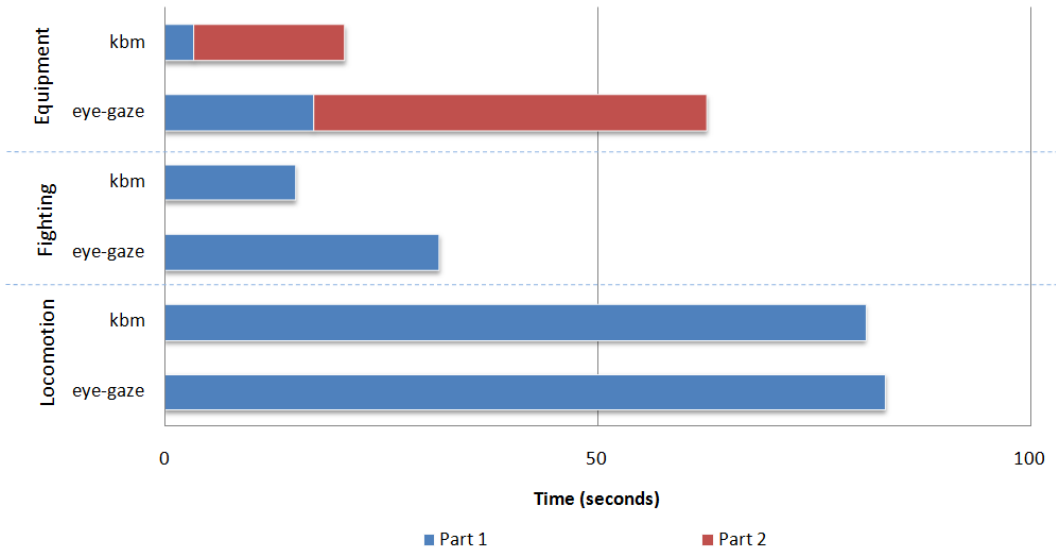


Figure 5.19: Evaluation Study 2: Task completion times and keyboard/mouse to gaze ratio. The ‘Equipment’ task requires two different interaction techniques: clicking (Part 1 on the chart) and dragging (Part 2 on the chart).

the subjective evaluation, 7 of the 10 participants said controlling the rate of turn of the character was especially difficult in the gaze condition. Fine control of changes in direction was said by some participants to be much easier with the keyboard than with gaze. The other control issue reported by 3 participants was the difficulty in starting and stopping movement quickly in the gaze condition (by glancing over the bottom edge of the screen). Also searching for a type of monster required reading the labels over the heads of characters as they appeared on screen. If these appeared on the right or left sides, reading the labels would cause unintentional turns in that direction.

5.5 Evaluation Study 2: Gaze Interaction in World of Warcraft

Another participant referred to the problem of feedback where it was difficult to see whether the characters had turned far enough when looking at the bottom left or right hand corners of the screen. Three participants rated gaze control of locomotion to be easier than keyboard and mouse as there was no need to keep pressing a key to move.

There was still observed instances of the ‘turn overshoot’ error however, these did not result in significant recovery time loss but they did lead to more deviations from the centre of a forward path movement. Another observed gaze specific error was ‘distraction’ where another character took the participant’s visual attention to part of the screen which caused the own character to turn. This also caused path deviation, which had to be corrected.

Fighting Task

All participants completed the task in both conditions. The data from the fighting task can be seen in Table 5.3 with the fight duration used as the measure. Here, the duration was measured from when the character first engaged the monster until the monster died. The gaze fight lasted twice as long as the fight in the keyboard/mouse condition; keyboard/mouse to gaze ratio of 1:2.1.

In this simplified fighting task, the main requirement was to click the spell icon continuously to cast as many spells as possible. In the subjective evaluation, 5 of the 10 subjects considered the size and location of the spell buttons to be a major factor with the difficulty of the task in the gaze condition. The magnifier was not used by any of the participants. When asked whether they considered using this to select the spell, one participant said that the number of steps required in order to select the magnifier, drop it and then select the spell was too distracting from the action during the fight. This is an important indicator in the design of gaze interaction techniques for tasks that involve interaction with other characters. Two participants said it was difficult to control the character during the fight as it was not possible to do multiple actions at the same time, such as moving and casting spells. Another participant explained that it was sometimes difficult to engage a fight with a monster as they would occasionally move around. Another participant noted how difficult it was to not look at the battle while they needed to keep looking at the spell button in the tool bar at the bottom of the screen.

5.5 Evaluation Study 2: Gaze Interaction in World of Warcraft

This task required rapid changing between modes to move, engage the character with a right click and then to cast spells with a left click. Three participants noted that they found changing modes quickly by glancing off screen difficult. However, they also suggested that this may be improved with more practice.

Equipment Task

All participants completed the task in both conditions. The results are shown in Table 5.3 with the task separated into 2 parts; opening of the pouch window and the character sheet window (part 1), and moving each of the four items from the pouch to the character sheet (part 2). The icons to open the two windows were situated at the edge of the screen and some participants found selecting these by gaze difficult due to the tracking accuracy being poorer nearer to the screen edge.

The keyboard/mouse to gaze performance ratios for the first and second parts of the tasks were 1:5 and 1:2.6 respectively. This gives a measure of the difference in difficulty between the two parts. Some participants used the magnifier in the gaze condition but only after they had tried to select the targets unaided, this resulted in long task times. Also dropping the magnifier at the bottom of the screen meant that half of the magnifier was clipped, which could, in some cases, obscure the enlarged view of the target icon.

In the subjective evaluation, opinion was divided between those who thought the task was easy to complete and those who found the first part (opening the equipment windows) and consequently the whole task difficult. 4 of the 10 participants rated the ease of the task completion with gaze as being either as easy as or easier than with mouse and keyboard. There may be an order effect as this task always followed the fighting task in both the gaze and the keyboard and mouse conditions, and so may have been considered easier overall.

5.5.4 Evaluation Study 1 and 2 Discussion

Both studies have demonstrated the feasibility of gaze control of MMORPGs in as much that all participants were able to complete all of the tasks. The ratio of task time using gaze to the time taken to complete the same task with keyboard and mouse was used as the main performance indicator. This allows some comparisons to be made between

5.5 Evaluation Study 2: Gaze Interaction in World of Warcraft

games and virtual environments provided the limits of similarities between the games and their tasks are recognized.

The first evaluation study, carried out with Second Life suggested that if the causes of identified problems in controlling locomotion could be designed out, then a performance ratio of keyboard/mouse to gaze in the region of 1 : 1.2 could be expected. Participants from the second evaluation study were able to achieve performance ratios of 1 : 1.1 or better. The main problem with gaze control of locomotion is the lack of fine control over the rate of turn of the character. To some extent, this is a problem with the game client as well as with gaze, and there have been some discussions on forums about the need for better rate of turn control when using keyboard and mouse control with the World of Warcraft client. Further, it is recognised that the task given to participants was restricted to moving in a fixed path, and not moving in response to dynamic events in the game.

The equipment and the fighting task both experienced the familiar problem in the difficulty of selecting small targets using gaze. The version of the magnifier used as a means of overcoming accuracy problems apparent from the first experiment was not an effective solution to this. Some of these problems may be attributed to specific implementation issues and some to the lack of training the participants had with the interaction technique. However, the main problem appeared to be the means of invoking the magnifier; picking it up, moving it, dropping it, and clicking through it were too distracting and time consuming for it to be effective in a time-constrained game playing situation. An alternative means of using the magnifier needs to be found, or an alternative solution altogether to the accuracy issue is needed.

The equipment changing task shared some similarities with the appearance changing task in Second Life. That experiment suggested that if the accuracy issues with gaze selection could be resolved then a performance ratio of 1 : 2.5 could be expected. The part of the equipment task involving object selection away from the edge of the screen in these trials had a keyboard/mouse to gaze ratio of 1 : 2.6. The similarity in these ratios gives encouragement to the idea that gaze performance across games can be quantified using the ratio as a metric, and that there is some consistency between similar types of task.

One task that was not repeated from the original pilot study is the communication task. The reason for this is because we wanted to evaluate the G9 predictive text

interaction technique against similar methods found in Subsection 3.1.3. Therefore, this remains as future work to be completed.

5.6 Discussion and Conclusions

This Chapter has used a bottom-up approach to provide information on interaction technique requirements, benchmarking and basic usability data. A software device (Snap Clutch), was created as a middle-ware application that translates gaze input into mouse and keyboard events that can be sent to any game that does not use DirectInput or is restricted by software firewall. Snap Clutch' extensible framework allows different interaction techniques to be implemented as interaction or task modes that can be rapidly switched between by performing off-screen glances. Several interaction modes are available that allow for simple mouse emulation and novel interaction modes that allow predictive text entry or locomotion in a game. The range of techniques can be expanded further by implementing variants based on differing levels of complexity, accuracy and temporal requirements, and spatial positioning. These components can also describe to us what is required from the user and hardware in order for them to be used correctly.

Two evaluation studies were conducted in order to evaluate the implemented gaze interaction techniques and benchmark them against keyboard and mouse input. Both have been able to demonstrate that it is feasible to carry simple locomotion, fighting and equipment manipulation tasks using gaze alone in Second Life and World of Warcraft. The first evaluation study showed the performance differences between gaze and keyboard/mouse interaction using task time ratios for similar types of task. These values were as expected, when a similar task was performed within the second evaluation study. The study also highlighted the limitations of the current approach to using gaze for time-constrained interaction with World of Warcraft as an example of an MMORPG. Suggesting that certain tasks require a more lightweight and rapid gaze interaction technique that allows the user to maintain their attention on the centre of the screen.

The following Chapter involves a top-down analysis by examining game tasks as a means of identifying interaction technique requirements. The aim of this is to match game tasks to candidate gaze interaction techniques.

Chapter deliverables:

- *A collection of gaze interaction techniques developed using iterative design and testing and a means to create variants of them.*
- *A taxonomy of gaze interaction modification properties.*
- *An extensible software architecture that is available for free public download.*
- *Evidence from two evaluation studies, suggesting that a basic level of game-play is possible in immersive virtual games using only eye-gaze by able-bodied participants.*

Chapter 6

Game Task Analysis

The aim of this Chapter is to examine in-game task characteristics and consider how these can be used to help select gaze interaction techniques. This results in a method of task analysis being developed that is based upon expert game play and is applied to an MMORPG via an evaluation study. Existing task modelling methods are explored and two approaches to selecting interaction techniques based on game tasks are examined however, both suggest that a broader approach is required.

Parts of this Chapter can be found within the following conference publication. Some portions are reproduced as are but much is rewritten to suit the flow of the Chapter and the thesis. All authors have made essential contributions and a brief note on my contribution personal is shown here.

Vickers, S.; Istance, H. & Smalley, M. **EyeGuitar: Making Rhythm Based Music Video Games Accessible Using Only Eye Movements**. *Proceedings of the 2010 International Conference on Advances in Computer Entertainment Technology: ACE '10*, 2010

Matthew Smalley developed the interaction technique under my supervision. The experiment was coordinated by myself and I wrote the full version of the paper equally with Howell Istance.

6.1 Typical Tasks in an MMORPG

The previous two chapters have considered that virtual environment tasks consist of: locomotion and camera control, object manipulation, application control and communication. Using this as a starting point, a self-study analysis of an MMORPG was conducted in order to verify that these tasks applied and if any new tasks should also be considered. Taking the game Guild Wars as the MMORPG, two sessions of open game play were analysed two weeks apart. The first session was recorded as a beginner to the game with the second session as a novice. In between the two sessions, there was daily game play of 30 minutes to an hour so by the time of the second session a novice level of game play was possible. The first session began after a set of initial in-game tutorials lasting around 20 minutes followed by 15 minutes of further game play. The second session began after an initial warm-up period of 15 minutes. Each session was recorded using video screen capture and lasted 15 minutes.

The session videos were annotated into tasks using Elan. The tasks fell into the following categories or domains:

- Windows control: Leaving the game to manipulate the Windows environment
- Application control: Altering in game settings
- Object manipulation: Interaction with an object in the game
- Locomotion: Walking, running within the game
- Free gaze: Looking around/thinking/waiting but not performing a specific task
- Communication: Interacting with other players or monsters (including fighting)
- Camera control: Altering of the camera angle

The time spent performing each type of task can be seen in Table 6.1 and Figure 6.1 with the user as a novice shown on the left, and as a novice on the right. Interestingly, the task domains of object manipulation, free gaze, and application control have remained very similar between the sessions. Locomotion and communication are the most frequent tasks in the game, and in the novice play most of the time is spent running or walking about, perhaps exploring the world. As the user became experienced, there was more interaction with monsters and other players shown by an increase in

6.1 Typical Tasks in an MMORPG

Table 6.1: Task breakdown of open play in an MMORPG over two sessions

	Session 1			Session 2		
<i>Task Domain</i>	<i>Count</i>	<i>%</i>	<i>Duration (s)</i>	<i>Count</i>	<i>%</i>	<i>Duration (s)</i>
<i>Application Control</i>	11	12.3	113.9	9	10	90
<i>Camera Control</i>	22	9.6	87.9	35	15.9	142.4
<i>Communication</i>	24	17.6	161.7	43	33.1	296.4
<i>Free Gaze</i>	14	11.2	104	9	11.3	101.6
<i>Locomotion</i>	51	46.2	425	35	26.1	233.8
<i>Object Manipulation</i>	8	2.7	24.4	14	3.6	32.3
<i>Windows Control</i>	1	0.4	3.5	0	0	0
<i>Total</i>	131		920.4	136		896.5

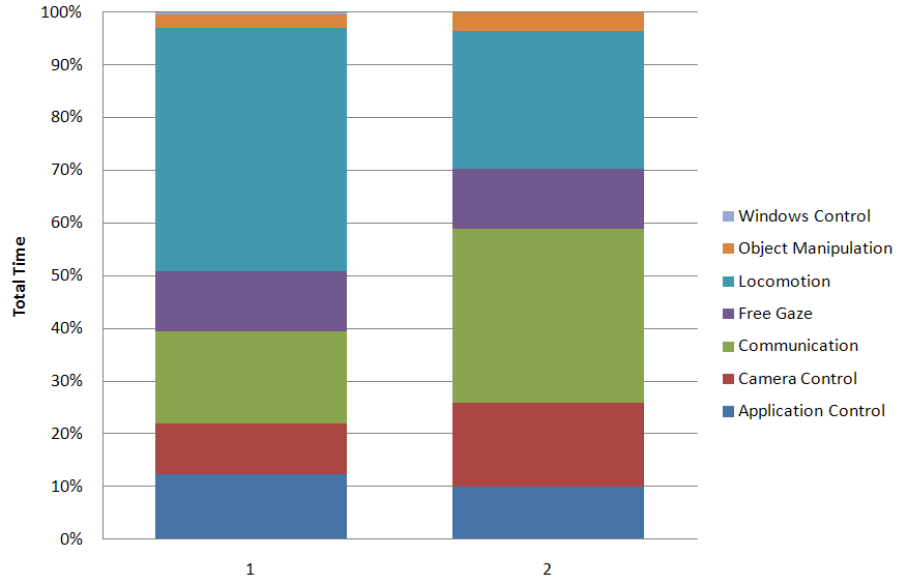


Figure 6.1: Task breakdown of open play in an MMORPG over two sessions

the communication and a reduced amount of locomotion, i.e. less exploration and more action.

Of all the tasks identified, locomotion appears as the most used in terms of frequency and as a percentage of overall time. It is also a task that has been investigated by researchers of immersive virtual reality environments (see Fajen and Warren, 2003;

Mackinlay et al., 1990; Ruddle and Lessels, 2006; Templeman et al., 1999).

6.2 A Methodology for Selecting Interaction Techniques based on Game Tasks

The methodology used here begins with establishing the requirements that eye-gaze based interaction techniques must satisfy. This involves task analysis of expert performance using keyboard and mouse when completing a series of typical game playing tasks. Following this, the tasks and interaction are modelled. Finally, candidate interaction techniques are selected based on the results from the task analysis and model.

6.2.1 Task Analysis

During the task analysis process the following data is collected simultaneously:

- Gaze position
- Think aloud and retrospective think-aloud protocols
- Video capture of the game window
- Logging of low level event data

The eye-gaze position provides us with information on where the user's attention is lying during a particular task. As non-command based techniques require natural, almost unintended control of objects by the user, it is important to understand where users naturally look when playing the game. This is crucial as in order to design or select a technique that feels natural to use it is important that it allows the users to focus their attention on what is happening on screen. The think-aloud protocol identifies what the user's intention is when performing a particular task and the retrospective protocol what they thought when watching their performance afterwards. The low level event data tells us what input events are required for each task performed and in what combination they are used.

To evaluate this approach a study was conducted containing gaming experts. In becoming experts, this user groups would have had to discover the most efficient methods of playing the game.

6.2.2 Evaluation Study 3A: Task Analysis of an MMORPG

The purpose of this evaluation study is to examine expert game play of standard tasks in an MMORPG.

Participants and Configuration

The study was carried out using the MMORPG World of Warcraft. An event logging application was developed in C# that captures all mouse and keyboard input events and writes them to an XML file. Participant's eye movements were recorded using a Tobii X120 eye tracker along with Tobii Studio software. All applications were run from an Antec desktop computer with 3GB of RAM and 256MB of Video RAM. Output was sent to a 20 inch wide screen Samsung monitor set at a resolution of 1680px x 1050px.

Five computing students from De Montfort University were recruited as participants for the study, aged between 18 and 20 (mean = 19 years). They were all male and experts in playing World of Warcraft. Each participant was encouraged to 'think aloud' as they performed the task. Therefore, a video camera was located behind them recording the screen and also audio commentary made by the participant. After the task was completed the video was played back so that additional retrospective comments could be made.

Procedure

Completing quests is a usual way to progress in a game like World of Warcraft with quests being collected from Non-Player Characters (NPC's). The standard task for this study was a quest where the player was asked to seek out a particular NPC and collect a quest from her. Once found, the NPC gave instructions that would guide the player on what needed to be accomplished. The participants all started at the same location (see Figure 6.2) and had no prior knowledge of what was involved in the quest. The quest required the participant to search for and then kill five skeletons before returning to the quest giver. Each was represented by a 'hunter' character, which was created with a medium skill level of 20. All of the abilities that are associated with a level 20 hunter were added to the character and placed along the game action interface located at the bottom of the screen (see Figure 6.2). The players were expert World of Warcraft players and were familiar with the hunter class of character and how to use the

6.2 A Methodology for Selecting Interaction Techniques based on Game Tasks

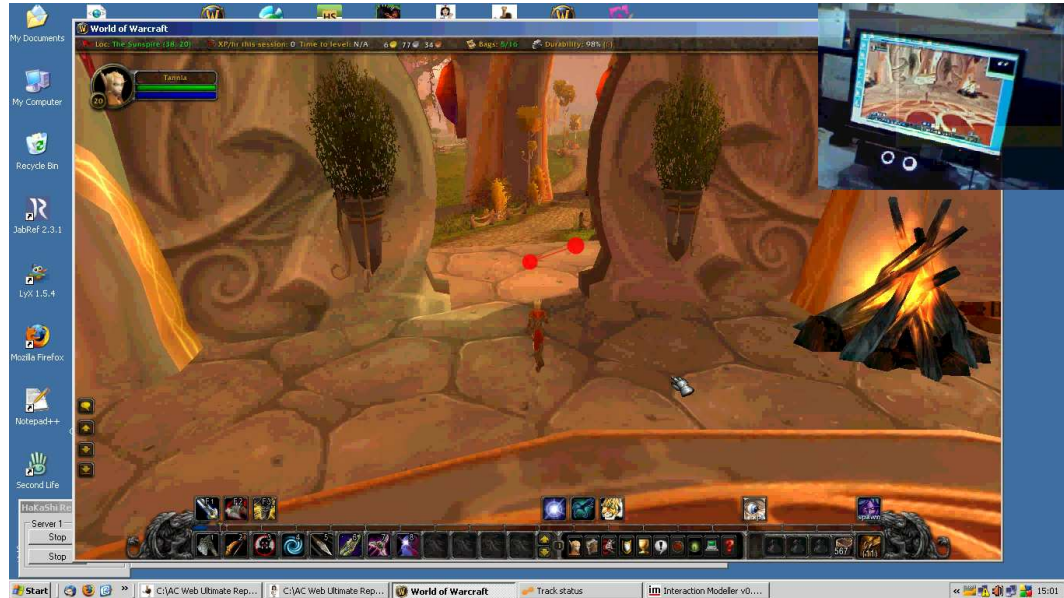


Figure 6.2: Evaluation Study 3A: This screen-shot is from the retrospective think-aloud video taken from the starting position of all participants. Part of the eye-gaze fixation scan path can be seen here by the two red circles in the centre of the screen. The video inlaid in the top right corner was taken by the external video camera that additionally captured all of the think-aloud audio from the participants.

characters special abilities. Each study lasted approximately 25 minutes. There were no restrictions placed on participants on how to control or manipulate their character.

Think Aloud Data

All verbs and nouns were collected from the transcripts and used to identify the tasks that were being performed. Table 6.2 shows a sample from one of the transcripts along with the task that was being performed at that time.

This process identified five key tasks that were performed:

- Search for NPC/mob
- Receive/complete quest
- Prepare to attack
- Fighting
- Looting

6.2 A Methodology for Selecting Interaction Techniques based on Game Tasks

Table 6.2: Evaluation Study 3A: A sample summary of the transcript taken from a participant during the experiment

<i>Time</i>	<i>Transcription</i>	<i>Task being performed</i>
00:08.3	‘Running’	Locomotion
00:15.2	‘Running’	Locomotion
00:17.6	‘Looking for Quest Giver’	Searching
00:22.7	‘Receiving quest’	Quest collection
00:26.5	‘Running’	Locomotion
00:35.2	‘Found target’	Searching
00:38.2	‘Blasting target’	Fighting
00:50.7	‘Searching for silver’	Loot corpse

Eye Tracking / User Attention Data

The Tobii Studio eye tracking analysis software was used to annotate the captured screen footage into the five key task categories. This produced two sets of data: eye tracking data for each individual absolute task and a time-line of tasks.

Using the annotations it is possible to show eye gaze data just for those task periods. During the ‘Search for NPC/mob’ task much of the participants gaze is focused on and around the avatar, see Figure 6.3. This is due to the participants gaze scan path moving from the avatar in the centre - out to the left or right - back to the avatar - and then to the opposite side, indicating a horizontal ‘scanning’ eye gesture. This produced a ‘butterfly’ visualisation with very soft edges leading from the dark red to the outer semi-transparent green. The fixation concentration is also distributed over a large area. This differs from the visualisation for the other absolute tasks (see Figures 6.3, 6.4 and 6.5), in which the fixation concentrations are less distributed and the areas around the avatars have harder edges indicating that the participant was looking at the avatar but not scanning around it. Note, that these Figures show all of the participants gaze data combined and not individual gaze data. The visualisations also shows how important some of the interface elements on the screen are. In the top left corner of the screen there is player and mob’s health information and in the bottom left is the button interface for performing spells. The gaze scan path for the fighting absolute task

6.2 A Methodology for Selecting Interaction Techniques based on Game Tasks

was performed by the participants gaze moving between these three Areas Of Interest (AOI) with little or no movement to other areas of the screen.



Figure 6.3: Evaluation Study 3A: Heat-map illustrating gaze behaviour of all participants for ‘Search for NPC/mob’ (left) and ‘Receive/complete quest’ (right) tasks.



Figure 6.4: Evaluation Study 3A: Heat-map illustrating gaze behaviour of all participants for ‘Prepare to attack’ (left) and ‘Fighting’ (right) tasks.



Figure 6.5: Evaluation Study 3A: Heat-map illustrating gaze behaviour of all participants for the ‘Looting’ task

6.2 A Methodology for Selecting Interaction Techniques based on Game Tasks

Task Time-line

Examining the annotation, a pattern in the overall task construction becomes apparent. Figure 6.6, shows the task structure and order that was typical amongst four of the five participants. Although, they spent different lengths of time performing each task, they were mostly performed in the same order. For example, ‘Search for NPC/mob’ would always be followed by a ‘Prepare for fight’, then ‘Fighting’, then ‘Looting’ before repeating all over again. The time-line for the fifth person can be seen in Figure 6.7, here he has shown signs of being less experienced than the other players by only performing one instance of the absolute task ‘Prepare for fight’. However, there still remains a pattern to the order of tasks that was performed.

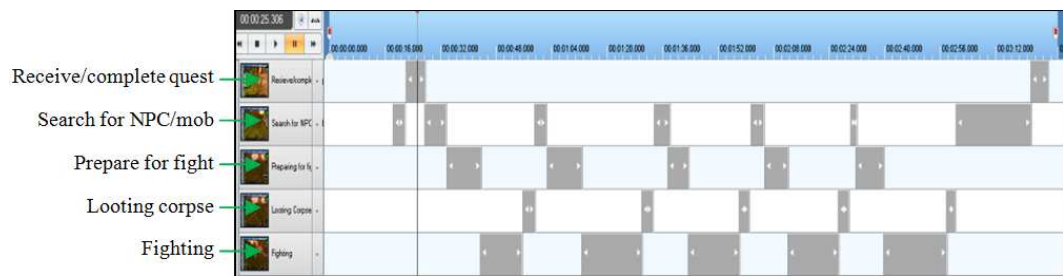


Figure 6.6: Evaluation Study 3A: Time-line of absolute task annotations for one participant



Figure 6.7: Evaluation Study 3A: Time-line of absolute task annotations for a second participant

Low Level Event Data

A MATLAB script was developed to parse the XML file containing the data to produce a time-line visualisation or event signature in a way similar to that used by ExperiScope

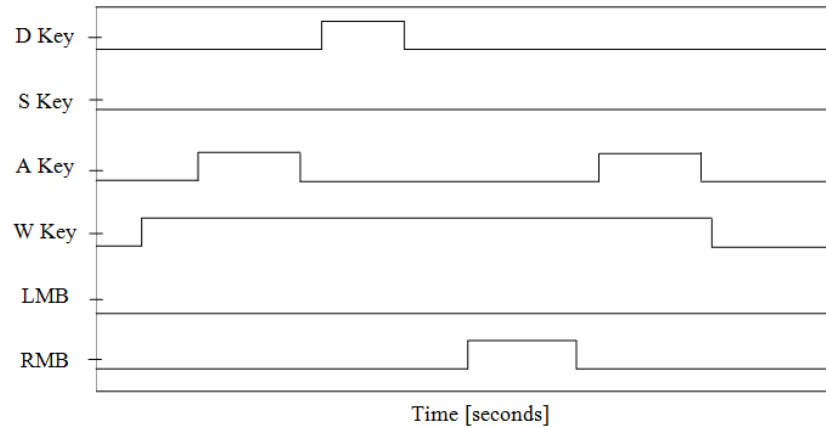


Figure 6.8: An example of the event signature visualisation

(see Guimbretiere et al., 2007). Each key event and mouse event is represented in the form of a horizontal on/off square wave that reflects the state of that event at that particular moment in time. It gives the ability to directly see what events are occurring at what time and if they are occurring in parallel to other events, see Figure 6.8.

Unfortunately, an issue with the implementation meant that many of the recorded events were shown to be out of synchronisation, for example, a key up event occurring before the key down event. It was discovered that this was due to a combination of keyboard and mouse buffering of events and the event logging software operating under normal process priority. The logging software was modified to use real-time process priority so as to capture events more accurately. The captured data from the evaluation study did provide event data types but due to this issue it could not be here used for visualisation.

6.2.3 Task Modelling of the Locomotion Tasks

There are many methods that have been used to represent and describe tasks and interaction¹, such as:

- GOMS (Goals, Operators, Methods, and Selection rules)
- BNF (Backus-Naur Form)
- UAN (User Action Notation)

¹For detailed explanations and discussions on GOMS, BNF, UAN and HTA, see Dix et al. (1998)

- KLM (Keystroke Level Model)
- HTA (Hierarchical Task Analysis)
- ExperiScope (see Guimbretiere et al., 2007)

Such approaches typically use visualisation and/or linguistic descriptions of what is taking place or what the user is intending to do. Taking a simple locomotion task in World of Warcraft as an example we can apply these methods. The locomotion task is to simply move the character using the keyboard freely for five seconds. Snap Clutch and the transparent overlay interaction technique is also analysed to provide an additional comparison.

GOMS

GOMS, models tasks in terms of sets of Goals, Operators, Methods and Selection rules (see Card et al., 1983).

- Goals: These are the user's objectives and describe what is to be achieved.
- Operators: These are the basic actions that a user must perform. They can affect the system and the user's state.
- Methods: There may be a number of ways to complete a goal or sub goal. Each one represented as a method.
- Selection: This defines the rules to which method is used and can depend upon the state of the system or the state of the user.

The result of a GOMS study for the keyboard can be seen in Figure 6.9 and for the transparent overlay in Figure 6.10.

GOMS is a widely used tool that is simple to apply and can be performed in a relatively short amount of time. Additionally, the hierarchical task structuring that is required performs a function that is comparable to that of HTA. However, it is difficult to add any kind of cognitive process, for example, 'check mini-map', 'stay on path', 'avoid monster' and so on. Representing this step which, occurs in parallel to the actual task is not possible using the current GOMS methodology family. Of course, the GOMS view of the task is abstract and not actual so there is an argument not to

6.2 A Methodology for Selecting Interaction Techniques based on Game Tasks

```

GOAL: WALK-TO-WAYPOINT-AND-RETURN-TO-STARTING-POINT
  GOAL: WALK-TO-FIRST-WAY-POINT
    [select GOAL: USE-W,A,S,D-KEYS-METHOD      repeat until way point reached
      • PRESS-W,A,S,D-KEYS
    GOAL: USE-ARROW-KEYS-METHOD                  repeat until way point reached
      • PRESS-CURSOR-KEYS]
  GOAL: WALK-TO-STARTING-POINT
    [select GOAL: USE-W,A,S,D-KEYS-METHOD      repeat until way point reached
      • PRESS-W,A,S,D-KEYS
    GOAL: USE-ARROW-KEYS-METHOD                  repeat until way point reached
      • PRESS-CURSOR-KEYS]

```

Figure 6.9: GOMS description of a locomotion task performed in World of Warcraft using a keyboard.

```

GOAL: WALK-TO-WAYPOINT-AND-RETURN-TO-STARTING-POINT
  GOAL: WALK-TO-FIRST-WAY-POINT
    GOAL: USE-LOCOMOTION-MODE
      • GLANCE-OFFSCREEN
    GOAL: USE-OVERLAY-ZONES                        repeat until way point reached
      • GLANCE-INTO-OVERLAY-ZONES
  GOAL: WALK-TO-STARTING-POINT
    GOAL: USE-OVERLAY-ZONES                        repeat until way point reached
      • GLANCE-INTO-OVERLAY-ZONES
  GOAL: STOP-AVATAR-MOVING
    [select GOAL: USE-LOOK-AROUND
      • GLANCE-OFFSCREEN
    GOAL: USE-LEFT/RIGHT
      • GLANCE-OFFSCREEN]

```

Figure 6.10: GOMS description of a locomotion task performed in World of Warcraft using the transparent overlay eye-gaze interaction technique.

include such processes that occur. As the task is presented in an abstract form there is also no method to represent time or repeatability.

BNF

BNF is a widely used form of notation used in describing the syntax of computer programming languages and other formal languages. It is considered a production rule grammar as it uses rules that describe correct strings and sentences. In general, a production rule grammar consists of:

6.2 A Methodology for Selecting Interaction Techniques based on Game Tasks

- A set of terminal symbols (words used in the language; lower case)
- A set of non-terminal symbols (constructs that show the language structure; upper case)
- A starting symbol (the descriptive start of the rule)
- A meta symbol (+, &, | ::= and so on)
- Rules constructed using the above (e.g. Symbol ::= noun phrase + verb phrase)

The results of a BNF study for the keyboard can be seen in Figure 6.11 and for the transparent overlay in Figure 6.12.

walk-to-waypoint-and-return-to-starting point	::=	walk-to-first-waypoint + walk-to-start-point
walk-to-first-way-point	::=	walking
walk-to-start-point	::=	walking
walking	::=	PRESS-W-KEY PRESS-A-KEY PRESS-S-KEY PRESS-D-KEY

Figure 6.11: BNF representation of a locomotion task performed in World of Warcraft using a keyboard.

walk-to-waypoint-and-return-to-starting-point	::=	walk-to-first-waypoint + walk-to-start-point
walk-to-first-waypoint	::=	select-locomotion-mode + walking
walk-to-start-point	::=	select-locomotion-mode + walking
select-locomotion-mode	::=	GLANCE-OFF-SCREEN
walking	::=	position-gaze
position-gaze	::=	MOVE-GAZE + position-gaze

Figure 6.12: BNF representation of a locomotion task performed in World of Warcraft using the transparent overlay eye-gaze interaction technique.

Following Figure 6.12, the first line shows the main goal, ‘walk-to-waypoint-and-return-to-starting-point’ is defined as (or ‘::=’) ‘walk-to-first-waypoint’ and ‘walk-to-start-point’. These non-terminals can be seen as the next two lines with their associated definitions. The non-terminal ‘select-locomotion-mode’ is defined as terminal ‘GLANCE-OFF-SCREEN’. This is the first of only two terminals and represents the physical action of the user glancing off of the screen. The second terminal ‘MOVE-GAZE’ and is part of the ‘position-gaze’ non-terminal. This represents the user moving their gaze position around the screen.

6.2 A Methodology for Selecting Interaction Techniques based on Game Tasks

The BNF description can be analysed in several ways. Firstly, the more rules that an interface requires then the more complex it becomes. Although, rules can be represented in different forms making this an unreliable measure. A more robust measure may be to count the number of meta symbols used. Using this measure we could say that the BNF analysis using the keyboard is less complex than using the transparent overlay. Secondly, the BNF description identifies many of the basic actions that are used to perform a task thus, allowing a quick but crude assessment of task difficulty. However, BNF is lacking any cognitive representation and consistency both with language structure and use of command names and letters.

UAN

UAN (User Action Notation) describes the physical behaviour of the user and the interface as tasks are performed (see Hartson et al., 1990). It is based on HTA but is presented in a tabular form, an example is shown in Figure 6.13.

TASK: Move a file icon		
USER ACTIONS	INTERFACE FEEDBACK	INTERFACE STATE
~[file_icon] Mv	file_icon-!: file_icon!, ∀file_icon`!: file_icon`-!	Selected = file

Figure 6.13: Example of a UAN description.

The first column, ‘USER ACTIONS’, shows the physical actions of the user. The second column, ‘INTERFACE FEEDBACK’ describes the feedback from the system based on the users actions. The final column, ‘INTERFACE STATE’ describes the result of the users actions. The description is a very low level view of the task but it does show the close relationship between the users actions and their associated reactions from the system. The notation syntax is descriptive but requires some training in order to understand all of the symbols.

Describing the locomotion task in World of Warcraft using UAN is difficult as there is little feedback from the interface itself. There is feedback from the system that is shown by the avatar moving in the world. However, this is not interface feedback and so is ignored by the notation. Figure 6.14 shows the UAN description of the task performed using a keyboard and 6.15 shows the UAN description using the transparent overlay.

6.2 A Methodology for Selecting Interaction Techniques based on Game Tasks

TASK: Walk to waypoint and return to starting point		
USER ACTIONS	INTERFACE FEEDBACK	INTERFACE STATE
W _v OR A _v OR S _v OR D _v	-	-
W [^] OR A [^] OR S [^] OR D [^]	-	-

Figure 6.14: UAN description of a locomotion task performed in World of Warcraft using a keyboard.

TASK: Walk to waypoint and return to starting point		
USER ACTIONS	INTERFACE FEEDBACK	INTERFACE STATE
~[off_screen]	mode_indicator-! : mode_indicator!	selected = locomotion_mode
~[walk_ahead_region] OR ~ [walk_left_region] OR ~[walk_right_region] OR ~[walk_backwards_region]	-	-
~[off_screen]	mode_indicator-! : mode_indicator!	selected = look around mode

Figure 6.15: UAN description of a locomotion task performed in World of Warcraft using the transparent overlay eye-gaze interaction technique.

The transparent overlay description offers much more information on the task and interface. That is, the user glances off screen and a mode indicator appears (a green bar appearing at the corresponding screen edge) effectively changing into locomotion mode. The user then must look into one of four regions to move before making another off screen glance reverting back to look around mode. Although, the above descriptions show a low level view of the task they only show the task from the user-interface point of view; and this is the intention of UAN. What they do not show is any in-world feedback. Additionally, the task decomposition is showed to be very systematic and procedural when the task being performed although repeatable, is likely to be different every time.

ExperiScope

Guimbretiere et al. (2007) developed a tool for analysing interaction data. ExperiScope

6.2 A Methodology for Selecting Interaction Techniques based on Game Tasks

presents a visualization time-line of events that is created using aspects of KLM Card et al. (1983) and Buxton's three-state model Buxton (1990). Three elements of the interaction are captured:

- The atomic tasks as required by the interaction
- The current state of the input device
- The user's level of engagement

Oppose to KLM's textual representation of task actions, ExperiScope uses graphical symbols, see Figure 6.16.











Action	Symbol	Description
Mark	L 	User performs a structured gesture or a mark.
Gesture	G 	User performs an unstructured, free form gesture.
Point	P 	User aims the pointer at a specific position on the screen.
Cross	C 	User crosses a specific goal on the screen [1].
Timeout	T 	User dwells at the same location for some time.
Mental Preparation	M 	User pauses to think before acting.
Home	H 	User homes (moves hand(s) to acquire an input device)
Button Press	K.R 	User presses and releases a button or keyboard key [29]
Visual shift	V 	User shifts visual attention to a different area of the screen
2-Handed Input	2 	Coordination between pointing devices controlled by each hand.

Figure 6.16: ExperiScope Symbols used to represent task actions.

There are several advantages in using this approach: firstly, the symbols are simple and extensible and secondly, it is easy to identify patterns or strings of actions. An example of ExperiScope's visualization can be seen in Figure 6.17.

In this example the user has been asked to perform a 'connect the dots' task. This involves selecting a colour on a tool palette and connecting two dots using a rubber band style of interaction. The Command Area, located in the upper portion of the notation, represents command selection actions such as using menu bars, screen buttons and so

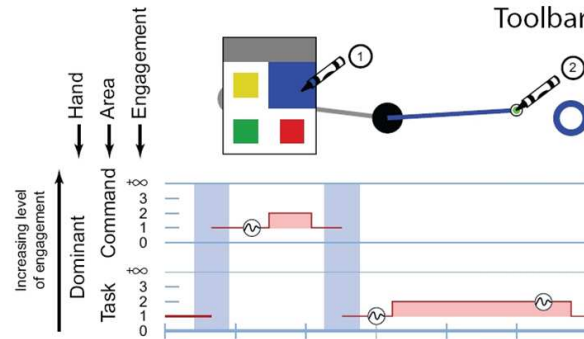


Figure 6.17: ExperiScope representation of connecting two dots using a tool palette. The interaction being performed is shown at the top and the ExperiScope visualization at the bottom.

on. The Task Area, located in the bottom portion of the notation represents the users work area such as, the drawing area of a drawing application. Transitions between the two areas are represented by bands of light blue. The time-line of events begins with the user's attention shifting to the command area (highlighted by the first blue band), they then point to the required colour and perform a click (shown by the input state changing from 1 to 2), the user then shifts attention to the drawing area (highlighted by the second blue band), they then point to the first dot and perform a drag to the second dot before releasing.

Figure 6.18 shows the ExperiScope visualisation of the locomotion task being performed using a keyboard and Figure 6.19 shows the visualisation using the transparent overlay. Due to the nature and length of this task only an example of how it would appear is shown, this is not actual data only typical data.

As this task requires only keyboard control and no interface components there is no information to add in the Command Area. Each key must be represented by a different Task Area row so as it state can be properly represented. It can be seen that the W key is in a down state for long periods of time with intermittent A and S key presses. The transparent overlay version of the task shown in Figure 6.19 makes use of the Command Area, as a gesture is required from the user that is independent of the game area. The additional Task Area's are to represent the different actions performed by the overlay/transparent buttons that are used to emulate keystrokes. This could be represented using one Task Area, but much of the interaction will be ignored and

6.2 A Methodology for Selecting Interaction Techniques based on Game Tasks

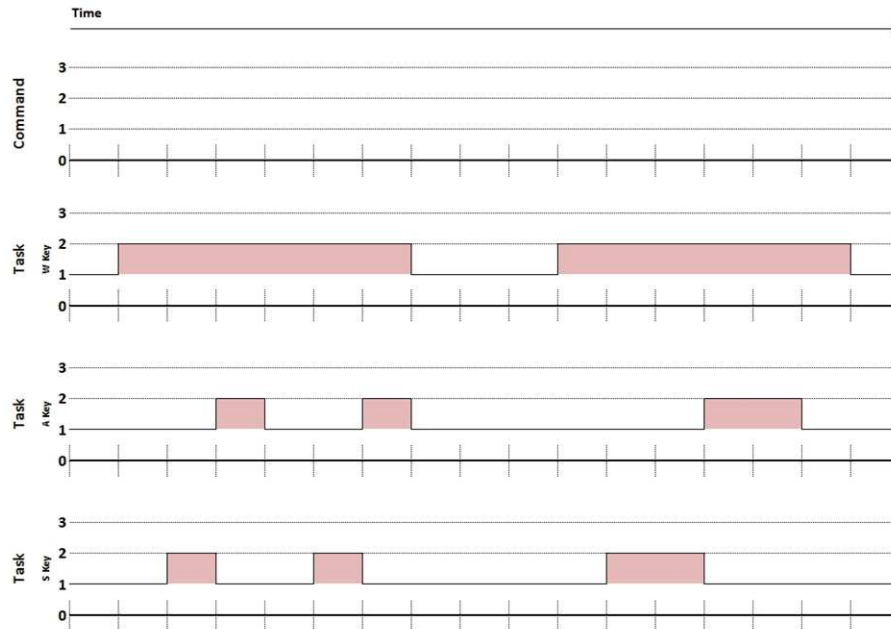


Figure 6.18: ExperiScope representation of the locomotion task performed in World of Warcraft using the keyboard.

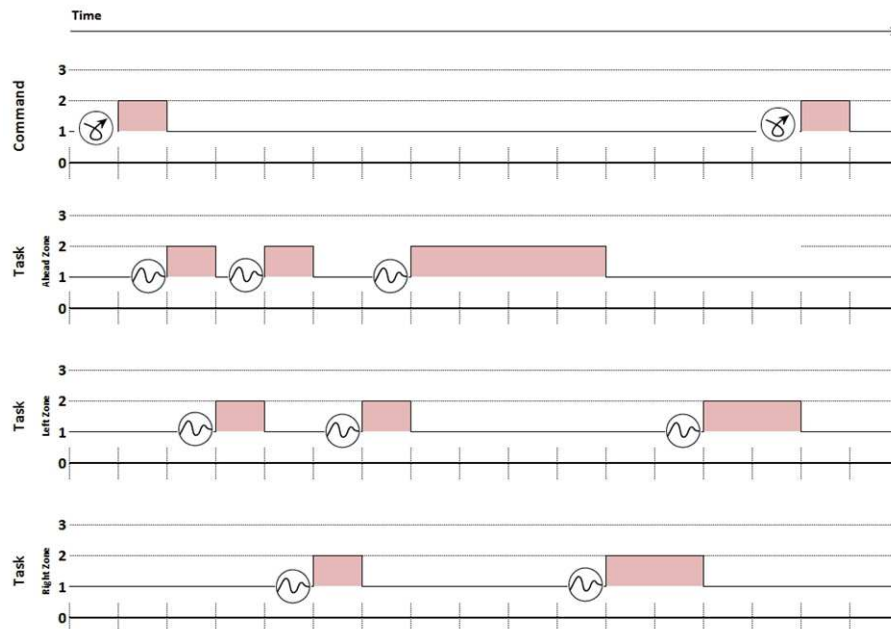


Figure 6.19: ExperiScope representation of the locomotion task performed in World of Warcraft using the transparent overlay eye-gaze interaction technique.

6.2 A Methodology for Selecting Interaction Techniques based on Game Tasks

there would be little visual representation of what is taking place. The separation of task and command works well for interaction in 2D desktop environments but does not represent the correct balance for 3D virtual environments.

All of the modelling methods applied offer insight into parts of the task or interaction but none of them offer a complete picture of what is occurring. Some methods are more abstract than others with many being designed for modelling interaction with 2D desktop style interfaces. The objective now is to build on the more useful parts of these methods, such that appropriate interaction techniques can either be suitably matched to the task or that sufficient information is provided to develop an entirely new interaction technique.

6.2.4 Object-Oriented Task Domains

Section 4.2 discussed the taxonomy of tasks found in virtual environments and the addition of a communication task domain. This domain classification can be extended further by adopting an object-oriented approach to modelling and representing tasks, Figure 6.20.

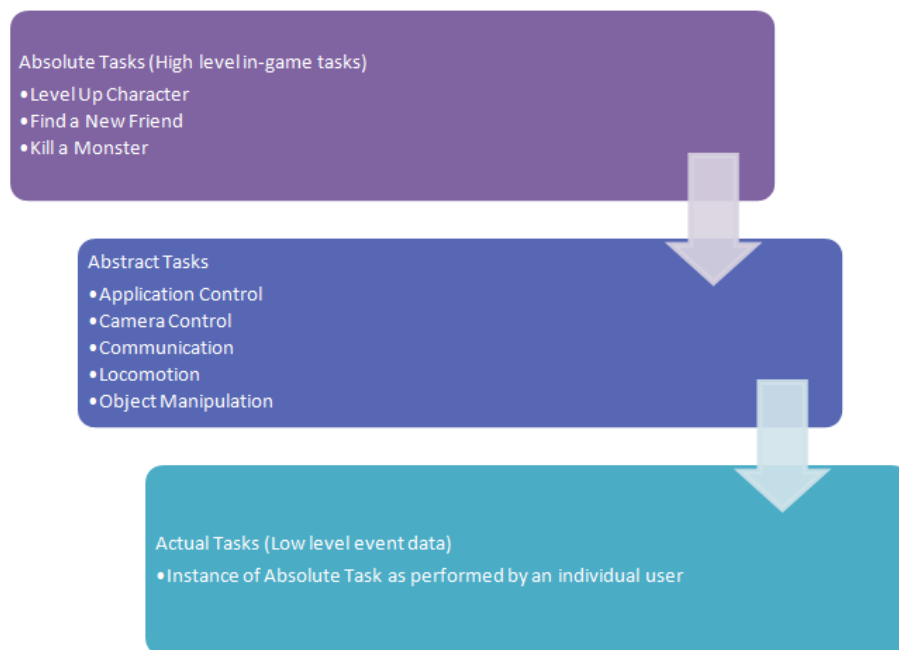


Figure 6.20: The three task classifications: absolute - abstract - actual

6.2 A Methodology for Selecting Interaction Techniques based on Game Tasks

Absolute tasks exist in the context of the actual game as a hierarchy of goals and sub-goals, in a similar structure to a plan found in HTA (see Annett, 2003). A higher level task can be something that is going to take longer such as ‘level up my character’ or something smaller such as ‘search for a new friend’ or ‘kill a monster’. Abstract tasks are low level tasks and are generic to virtual 3D graphics environments. So, absolute tasks require a combination of low level abstract tasks to be performed in order for them to be completed. For example, ‘kill a monster’ may require a combination of locomotion and application control tasks. When these abstract tasks are actually performed by a user, or instantiated, they then become actual tasks, these are similar to ‘methods’ as used in GOMS. The actual task may vary between users as there is often more than one way to perform each task. Additionally, each user will have a different level of skill and knowledge and this can be reflected in the actual task representation.

Modelling Game Tasks and Interaction

The absolute tasks can be represented in a notation form that uses elements of GOMS, BNF and UAN, see Table 6.3. In this example, (1) the task is declared as Absolute and given a name; (2) the Abstract tasks used within the task are declared; (3) the interaction required to perform the absolute task; (4) the notation closes.

Table 6.3: Absolute task notation structure

1	<i>Absolute SomeTask</i> {
2	<i>AbstractTasks</i> (<i>CAMERA</i> , <i>LOCOMOTION</i>),
3	<i>Interaction</i> { $\delta\infty(X,Y)$ }
4	};

The notation used to represent the interaction techniques is extensible with a basic set of notation found in Table 6.4. Taking a simple locomotion task as an example, the notation can be used to model as in Table 6.5:

Here, (1) the task is declared; (2) the task contains camera and locomotion abstract tasks; (3) start of interaction notation; (4) a Select statement is declared showing a choice of two Repeatable interactions that are separated by a ‘,’. The first indicates that either the W, A, S or D key should be depressed; (5) the second Repeatable interaction

6.2 A Methodology for Selecting Interaction Techniques based on Game Tasks

Table 6.4: Basic set of interaction notation

<i>Notation</i>	<i>Represents</i>
<i>LMB0</i>	Left mouse button click
<i>LMB1</i>	Left mouse down
<i>LMB2</i>	Left mouse up
<i>RMB0</i>	Right mouse button click
<i>RMB1</i>	Right mouse button down
<i>RMB2</i>	Right mouse button up
<i>Key_A0</i>	A key press
<i>Key_A1</i>	A key down
<i>Key_A2</i>	A key up
$\delta_{\infty}(X, Y)$	X, Y movement with infinite steps
$\delta_1(X, Y)$	X, Y movement with single steps
	OR operator
&&	AND operator
\rightarrow	next

Table 6.5: Absolute task notation for a locomotion task performed using keyboard and mouse

```

1 | Absolute WalkToWaypointAndReturnKBM{
2 |   AbstractTasks(CAMERA, LOCOMOTION),
3 |   Interaction{
4 |     [Select Repeatable(KeyW1 || KeyA1 || KeyS1 || KeyD1),
5 |     Repeatable((KeyW1 || KeyA1 || KeyS1 || KeyD1)
6 |     && (RMB1  $\rightarrow$   $\delta_{\infty}(X, Y)$ ))]
7 |   };

```

indicates that either the W, A, S or D key should be depressed in addition to (6) a right mouse button down and coordinate movement (a right mouse drag); (7) the notation closes.

The Snap Clutch ‘locomotion mode’ version of the notation appears as Table 6.6.

6.2 A Methodology for Selecting Interaction Techniques based on Game Tasks

Here, (1) the task is declared; (2) the task contains camera and locomotion abstract tasks; (3) start of interaction notation; (4) locomotion mode is selected by performing the corresponding glance then a NEXT operator indicating the flow of interaction; (5) a Repeatable statement of coordinate movement; (6) look around mode is selected by performing the corresponding glance; (7) the interaction notation closes; (8) the notation closes.

Table 6.6: Absolute task notation for a locomotion task performed using the ‘locomotion mode’ in Snap Clutch

1	<i>Absolute walkToWaypointAndReturnSC</i> {
2	<i>AbstractTasks</i> (CAMERA, LOCOMOTION),
3	<i>Interaction</i> {
4	<i>glanceModeLoco</i> →
5	<i>Repeatable</i> ($\delta\infty(X, Y)$) →
6	<i>glanceModeLook</i> }
7	};

Following the object-oriented approach, absolute tasks are instantiated when they are performed by a user making them become an actual task. In order to notate the actual task, elements of the actual task being performed are required. These can be gathered from video analysis or direct output from software capturing keyboard and mouse events in real-time. An example of how an actual task can be represented is shown in Table 6.7. Here, (1) the task is declared along with the abstract task it is instantiated from; (2) start of interaction notation; (3) each interaction step is now numbered and is given a time stamp. The first step is W key down; (4) A key down and then up; (5) D key down and then up; (6) right mouse button down and move to coordinates x=350, y=400 then release; (7) A key down and the up; (8) W key up; (9) notation closes. The information presented in the actual task can then be used to generate an ExperiScope style of visualization.

Although, this approach allows us to model game tasks and interaction it suffers from similar problems to it’s predecessors. For instance, the notation and it’s structure is complex making adoption undesirable. It also takes a great detail of time and space to notate the actual task model; the actual task model in Table 6.7 only represents one

6.2 A Methodology for Selecting Interaction Techniques based on Game Tasks

Table 6.7: Actual task notation model for the locomotion task performed using keyboard and mouse

1	<i>Actual walkToWaypointAndReturnKBM(User1){</i>
2	<i>Interaction{</i>
3	¹ <i>KeyW1</i> →
4	² <i>KeyA1</i> → ³ <i>KeyA2</i>
5	⁴ <i>KeyD1</i> → ⁵ <i>KeyD2</i>
6	⁶ <i>RMB1</i> → ⁷ $\delta\infty(350, 400)$ → ⁸ <i>RMB2</i>
7	⁹ <i>KeyA1</i> → ¹⁰ <i>KeyA2</i>
8	¹ <i>KeyW2</i> }
9	<i>};</i>

or two seconds of task time. A more convenient way to model the tasks and interaction is to perhaps use XML. This would also allow easy parsing of data to allow some form of visual display of the model or replay of the actual interaction. However, these attempts at modelling the tasks and interaction, is a reminder that the objective is to select an appropriate gaze interaction technique based on task properties and requirements. As useful as this descriptive method may be, it does not offer any real contribution in terms of our objective.

Using Task Requirements as Interaction Technique Properties

In contrast to using a descriptive model to represent game tasks and interaction, here we are trying to define interaction technique properties based upon the task requirements. Each abstract task is built of requirements that define it and separate it from other abstract tasks. These can be identified by reviewing data collected from task analysis. A matching interaction technique property can then be defined in terms of a simple boolean operator, as in Table 6.8 for instance. The expectation then may be that, the interaction techniques that are capable of answering the greater number of TRUE properties will be the ones most suitable for the task.

However, this process is flawed in that it assumes all task requirements are equal in their importance, frequency and so on. In order to consider these requirements it is necessary to conduct some form of weighting analysis. Analytical Hierarchical Process

6.2 A Methodology for Selecting Interaction Techniques based on Game Tasks

Table 6.8: Creating interaction technique properties from game task requirements. In this example, some important requirements of a locomotion task are matched with possible interaction technique properties.

<i>Locomotion Task Requirement</i>	<i>Interaction Technique Property</i>
Requires key or mouse events	Can perform the necessary range of key and/or mouse events?
Obtain feedback of current location	Can look at other objects without affecting interaction?
Vary rate of turn	Can vary the rate of turn?
Vary rate of speed	Can vary the rate of speed?
Start/stop movement at precise location	Can start/stop movement instantly and accurately?
Avoid obstacles on route	Can quickly change speed or direction?
Precisely position the avatar	Can finely control the avatar movement?

(AHP) is a method of Multi-Criteria Decision Making (MCDM) (see Satty, 1990). Methods of this kind take a subjective strength of preference approach rather than taking specific values of measurement. As preferences and objectives differ between users the result of the analysis will also change. The process is not intended to give a definitive answer to a problem but to suggest how different answers may fulfil their needs, requirements and their own understanding of the problem.

The hierarchy consists of the overall goal, the options or alternatives to achieving the goal and the criterion that relate the goal to the alternatives, see Figure 6.21.

In applying the method to game tasks and gaze based interaction techniques we consider that:

- Our **goal** is to select an interaction technique
- The **criterion** are the abstract task requirements and,
- The **alternatives** are the interaction techniques to choose from

The number of preferences made is dependent upon the number of criterion. Only six of what is considered to be the most important criterion are used in order to reduce

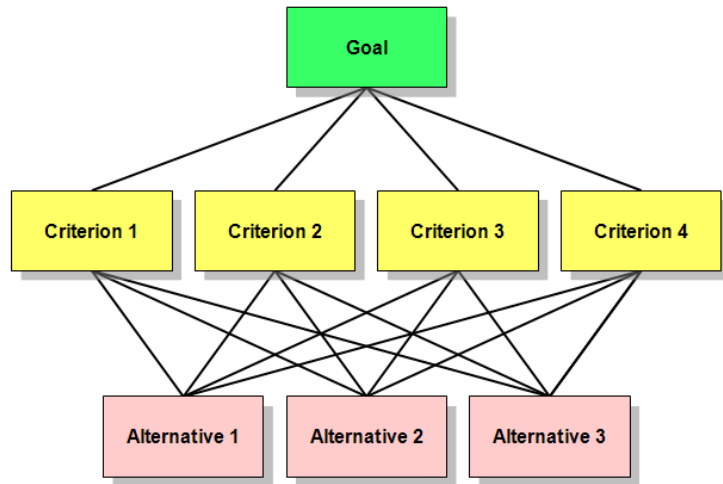


Figure 6.21: AHP Hierarchy

the possible number of comparisons. A pair-wise comparison with a scale of 1 to 9 is used to then determine the preference. When performing the comparison it is important to consider the problem in two ways; in terms of importance and also frequency. Thus, this accounts for two separate goals and therefore, two separate AHP analysis'.

However, the difficulty in this approach now lays in that there are advantages and disadvantages with each interaction technique. For instance, if two different interaction techniques both answered TRUE to the property 'Can vary the rate of turn?' it implies that they both fulfil this requirement equally. This is not the case and one technique may be vastly superior to the other when it comes to actually varying the rate of turn. Further, there is no consideration for the user or hardware being used. For instance, one technique may perform better than another if a more accurate eye tracker is being used or one person may prefer to use one interaction technique over another.

Both investigations into modelling tasks and interaction for the purpose of appropriate selection have shown to be too limited and difficult to use to be of practical value. The descriptive method is complicated and limited in its information provision and the mapping method based on requirements and properties does not consider enough about the hardware or the user. This all suggests that a more general and broader task modelling method is required.

6.2.5 Selecting an Appropriate Gaze Interaction Technique

All of the data collected using the task analysis method contributes to the selection or design of an appropriate eye gaze interaction technique. There are two key requirements that we can consider: user attention and input events. Summarising the task data from evaluation study 3A we can say the following (Table 6.9):

1. The *'Receive/complete quest'* task requires three different input events: moving the cursor; left and right mouse clicks. The two AOIs are the NPC that is first interacted with; reading of the quest text. This task has no time pressure associated with it. This suggests that there is a need for accurate pointing and selection thus, the candidate gaze interaction techniques for performing this task include mouse emulation, a magnifier or tool glass.
2. The *'Search for NPC/mob'* task requires four different input events: W, A, S and D keys. These is a requirement for these to be in a down state for different lengths of time as well as being used in combination with one another. The only AOI is a large butterfly pattern spreading across the horizon. This task has no time pressure associated with it but it is necessary to turn the character quickly when required (eliminating the use of dwell buttons). This suggests that there is a need to look at other objects without it affecting the task thus, the candidate gaze interaction techniques for performing this task include transparent overlay.
3. The *'Prepare for fight'* task requires two different input events: moving the cursor; left mouse click. The three AOIs are the health bars, the character that is about to be attacked, and the spell buttons at the bottom of the screen. This task has no time pressure associated with it. This suggests that there is a need for accurate pointing and selection, thus the candidate gaze interaction techniques for performing this task include mouse emulation, a magnifier or tool glass. However, as there are three distinct AOIs it may be possible to assign 3-stroke gestures to key bound macros.
4. The *'Looting corpse'* task requires three different input events: moving the cursor; left and right mouse clicks. The two AOIs are the corpse and the loot window. This task has no time pressure associated with it. This suggests that there is a need for accurate pointing and selection thus, the candidate gaze interaction

6.2 A Methodology for Selecting Interaction Techniques based on Game Tasks

techniques for performing this task include mouse emulation, a magnifier or tool glass.

5. The ‘*Fighting*’ task requires two different input events: moving the cursor; left mouse click. The three AOIs are the health bars, the central area where the fight is occurring and the spell buttons at the bottom of the screen. This task is time critical. This suggests that there is a need for accurate pointing and selection, thus the candidate gaze interaction techniques for performing this task include mouse emulation. As this is a time-critical task, the use of a magnifier or tool glass is not recommended as it was shown in evaluation study 2 (Section 5.5) that these techniques take too long in performing. However, as there are three distinct AOIs it may be possible to assign 3-stroke gestures to key bound macros.

Table 6.9: Evaluation Study 3A: Summary of the game tasks and candidate gaze interaction techniques

<i>Task</i>	<i>Input Events</i>	<i>AOI's</i>	<i>Candidate Tech- niques</i>
<i>Receive/complete quest</i>	Cursor point, left and right click	Quest text, NPC	Mouse emulation, magnifier/tool glass
<i>Search for NPC/mob</i>	W, A, S, D keys (some in combination)	Horizon	Transparent overlay
<i>Prepare for fight</i>	Cursor point, left click	Health bars, NPC/mob, spell buttons	Mouse emulation, magnifier/tool glass, 3-stroke gestures
<i>Looting corpse</i>	Cursor point, left and right click	Corpse, loot window	Mouse emulation, magnifier/tool glass
<i>Fighting</i>	Cursor point, left click	Health bars, action area, spell buttons	Mouse emulation, magnifier/tool glass, 3-stroke gestures

Expanding further on the task ‘search for NPC/mob’. This locomotion based task requires the user to be free to examine the scene before them with much of the attention

6.2 A Methodology for Selecting Interaction Techniques based on Game Tasks

around the avatar and horizon. The low level event data not only provides information on what key and mouse events are required to perform a particular type of task; it also show frequency, duration, as well as input events that are performed in parallel to one another. The input events required for this task are W, A, S and D key events. The W key is normally in a down state for the majority of the time but there is a requirement to apply A and D key events in parallel. A technique that fulfils these requirements is the transparent overlay. When used for a locomotion task, it generates a stream of W key down events that are sent to the target application. The user attention data shows that the butterfly visualization is created by the user scanning around and so they should be free to do so with no rotation of the avatar. However, if they were to look further past the butterfly region then it can be assumed that they want to rotate their avatar, and so the technique can stream A and D key down events accordingly, see Figure 6.22. The attention requirement has partly been fulfilled as the user is free to look at parts of



Figure 6.22: Position of an on screen overlay (not visible to the user) that allows for gaze based locomotion. The heat-map underneath show the natural gaze positions for one participant whilst navigating in World of Warcraft using keyboard and mouse.

the world around the avatar without movement of the avatar. Looking to the left and right of the avatar does provide some natural action but it is not always the case that you want to rotate and face the object that you are looking at. In the case of the other tasks, the attention is always focused on a small number of AOIs, such as the avatar, health meter and interface attack buttons. This suggests that some form of gesture

based interaction could be applied that use these AOIs as activation zones. This would allow the user to keep their attention on what is happening on screen. There would also be no issues with eye tracker accuracy as we would be eliminating the need to hit small interface buttons. However, the necessary input events are a combination of movement controls and mouse clicks, so there is a requirement for mapping the correct events to specific gestures.

6.2.6 Evaluation Study 3B: Comparing Event Signatures Between Gaze and Keyboard

Due to the issue previously with visualisation of the event signature, a further study was performed to compare the keyboard with the transparent overlay technique.

Participants and Configuration

The study was carried out using World of Warcraft. The logging application from evaluation study 3A was modified as described previously in order to capture events more accurately. A Tobii X120 eye tracker was used along with Snap Clutch software. All applications were run from an Antec desktop computer with 3GB of RAM and 256MB of Video RAM. Output was sent to a 20 inch wide screen Samsung monitor set at a resolution of 1680px x 1050px. The same five participants from evaluation study 3A took part in this study.

Procedure

Prior to the start of the study each participant was introduced to Snap Clutch and allowed to practice using the transparent overlay technique for several minutes. Once they were comfortable with the operation the study began. All participants started at the same location and were asked to navigate a long winding path until they reached a large fence at the end of the path. They were to perform this twice, first of all using the keyboard and then using gaze. During self-completion the path was navigated in 1 minute 48 seconds, it had many turns with no obstacles therefore, there would be little to gain in any kind of learning effect. The only experimental recording was made by the logging application.

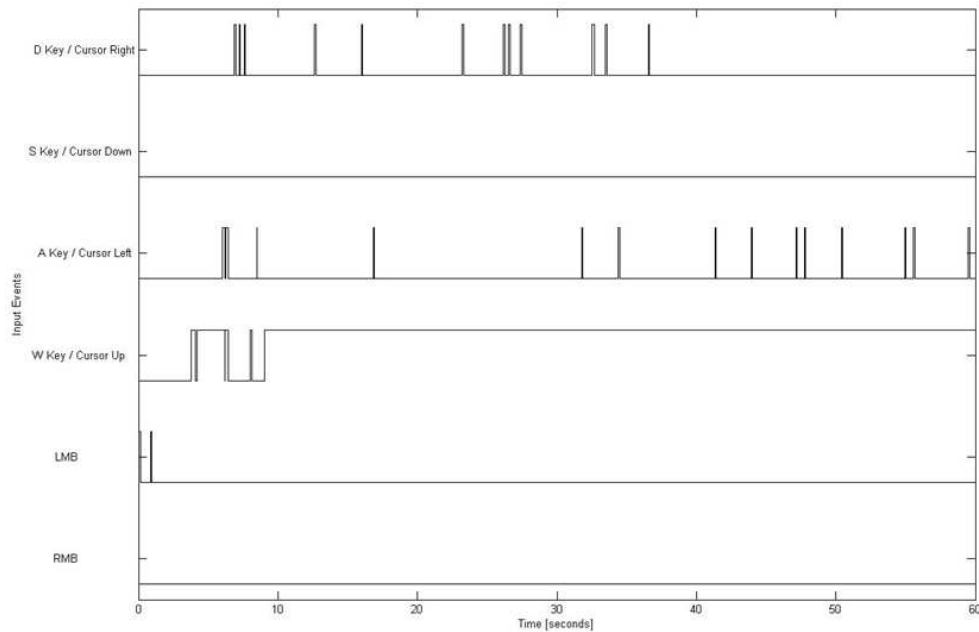


Figure 6.23: Evaluation Study 3B: Event signature for the first 60 seconds from one participant using the keyboard

Results and Analysis

The signature between participants was very similar when using gaze and when using keyboard. Figure 6.23 shows the event signature for the first 60 seconds from one of the participants. It can be seen that the user kept the W key down for much of the time but opted to use the A key and D key with quick presses to navigate their avatar along the path. The corresponding event visualisation for the first 60 seconds of the task using the gaze based interaction technique can be seen in Figure 6.24. Although not identical they are very similar in their appearance. The length of time that the A and D keys are pressed can be seen to be quite consistent with each other. This is an indication to the strength/variance of each left and right turn made.

6.3 EyeGuitar

The task analysis method presented at the start of this Chapter was also applied to a rhythm based music game similar to Guitar Hero¹. To play such games you must

¹The following is only a summary please see Vickers et al. (2010) for further explanations.

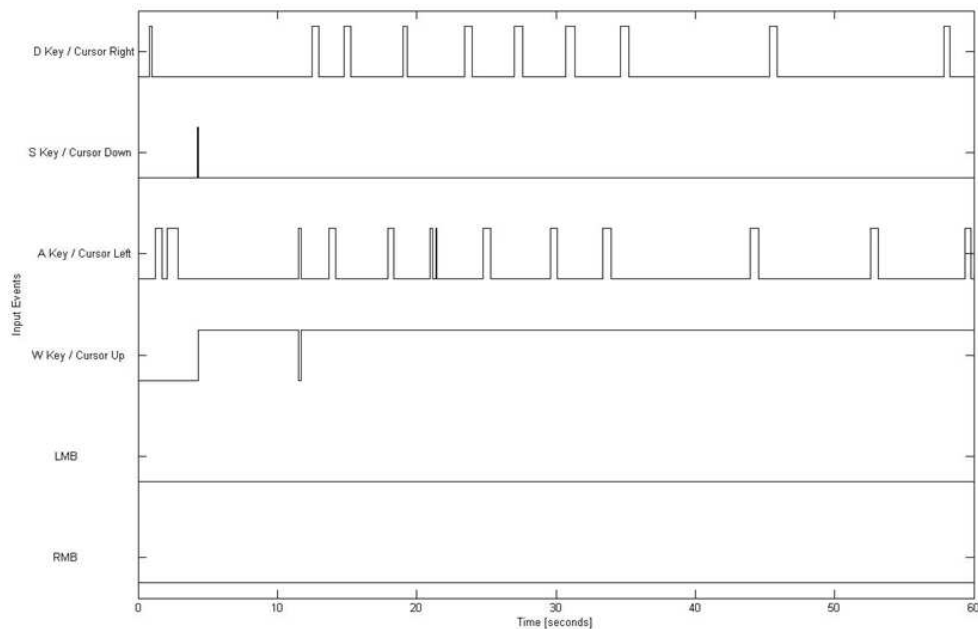


Figure 6.24: Evaluation Study 3B: Event signature for the first 60 seconds from one participant using eye-gaze (Snap Clutch and the transparent overlay)

press keys on virtual guitars in various combinations in time with the music. Although, this genre of game is designed to be played with a virtual guitar, gamers have made novel use of an ordinary computer keyboard to replicate the standard virtual guitar interaction technique. This is accomplished by turning the keyboard on it's side and holding it like a guitar. The left hand uses the first five Function keys as the fret buttons and the right hand uses Space bar to strum the notes.

Following the task analysis, a gaze interaction technique was developed that worked as follows. The user gazes at the notes as they travel down screen, see Figure 6.25. An on-screen paddle graphic moves below the note strum point in relation to the users gaze position and how far down the fretboard they are looking. The fretboard 'highway' shows notes coming from a distance along a 3D virtual road, creating a triangular shape. Therefore, the paddles horizontal movement is non-linear as it moves with the gaze position following the notes. The user is notified of the current note selection by the colour of a large on-screen bar that appears above the play area. The note is strummed automatically as it passes the strum bar.

This technique was evaluated and compared with able-bodied participants using

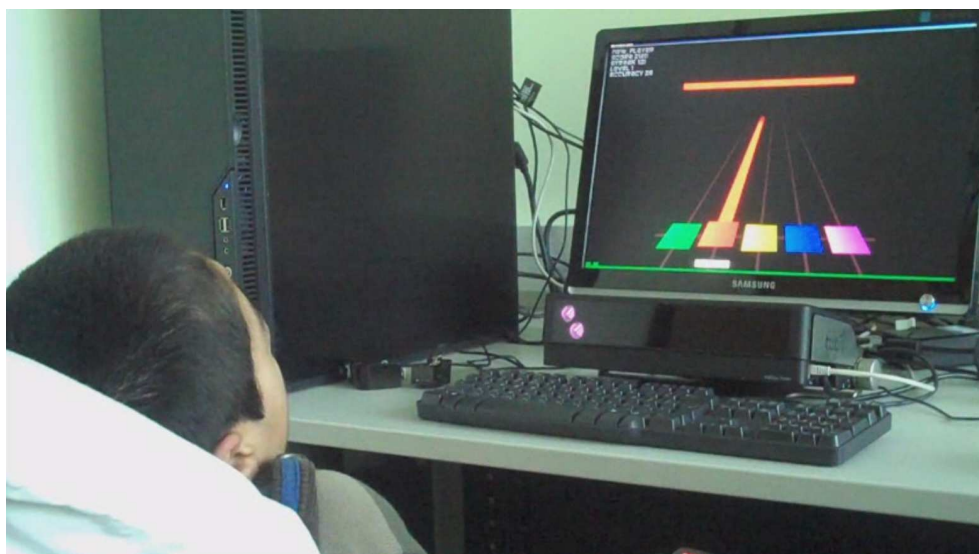


Figure 6.25: A participant suffering from Muscular Dystrophy playing EyeGuitar

the keyboard method and the results were comparable. The technique was then tested by three children (all aged 17) suffering from MD. All three struggled to get used to the technique and commented on how they were distracted by the paddle that moved along the bottom of the screen. This distraction problem meant that they were scoring poorly to begin with, with most of the notes being missed. However, this improved and the distraction of the paddle became less a problem and two were able to achieve more than 50% hit percentage. There was an issue with one participant unable to look at the notes coming down the left side of the screen when previously following notes on the right. This was due to the speed that the notes were travelling and the speed that the paddle moved. Although, this is a guitar based music game it was commented that it did not feel like they were playing or pretending to play the guitar. However, they did enjoy the game and believed that the interaction technique worked well although, being able to change the speed of the paddle may improve game play.

In terms of using the task analysis method as a means of selecting or developing an appropriate gaze interaction, it has been applied successfully. However, the interaction technique developed only suited the able-bodied participants and two of the MD children.

6.4 Concluding Summary

This Chapter has explored the possibility of using a top-down approach of task analysis followed by task modelling for aiding gaze interaction selection and creation. Task analysis and modelling has been a central part of human-computer interaction and this has resulted in many methods being developed. Elements of GOMS, UAN, Experiscope and so on were used in an attempt to model and describe game tasks, however the results from this suggested that a more broader approach was required. An approach to game task analysis was presented along with results from an evaluation study of expert game play in World of Warcraft. Here, gaze position was recorded to show user attention; think-aloud protocols were used to document user intention; and input events were captured to record the task event requirements from the input devices. By classifying the tasks performed into absolute-abstract-actual tasks, a time-line of events developed. When annotated, it allows for gaze attention heat maps to be created for each absolute task. Combining these together allowed for candidate interaction techniques to be matched in a basic form to game tasks. The methodology was also verified using an alternative genre of game similar to Guitar Hero.

The task analysis process provides essential information that establishes what the task requires from the interaction technique in order to be fulfilled. This may allow for a suitable interaction technique to be selected or a new one designed. However, the process does not consider enough about what the user's capabilities are. The assumption has been that a perfect calibration will be achieved and that each user will be capable of performing the interaction technique. The following Chapter focuses on extracting end-user requirements and provides the basis for dynamic accessibility.

Chapter deliverables:

- *An approach to task analysis and modelling that provides game task and user requirements of an interaction technique.*
- *Results from an evaluation study that matched candidate interaction techniques to typical game tasks found within an MMORPG.*
- *A visualisation model for displaying low level event data.*

Part III

User Evaluation

Chapter 7

Dynamic Accessibility

The experiments found in Chapter 5 and Chapter 6 have involved participants with no physical disabilities. The main conclusion that can be drawn from these studies is that a basic level of gameplay for a game such as World of Warcraft or Second Life can be achieved using only eye-gaze. Evaluation using able-bodied students is a useful first step for providing base-line performance data, fine-tuning tasks and techniques and eliminating software bugs. This helps in making the evaluation process with the intended end user much smoother.

This chapter presents the evolutionary process of a series of experiments aimed at enabling young people with physical disabilities to perform simple game related tasks in World of Warcraft based upon their individual specific needs. Additionally, it provides the basis for dynamic accessibility where the provision of accessibility can be adjusted rather than being considered constant.

7.1 Dynamic Accessibility

In this context, by dynamic accessibility we mean that levels of accessibility are unique to every user and thus needs to adapt depending upon the requirements of the user and the hardware that is available to them. The premise being that an individual with a specific disability; at a specific time of the day; in a specific physical and emotional state; may require a different level of accessibility to aid their communication. This may come in the form of temporarily adding or removing an AAC device or perhaps simplifying, automating and manipulating a process.

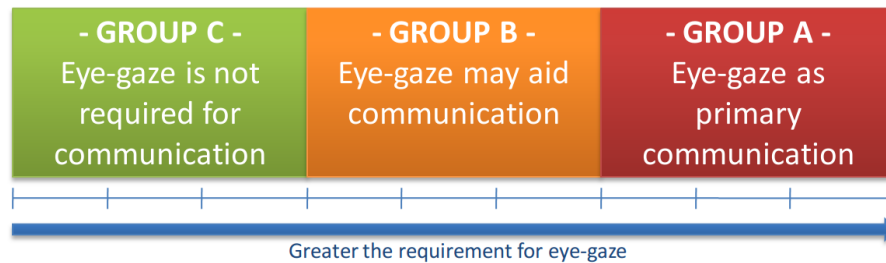


Figure 7.1: A scale of eye-gaze as a requirement for primary communication.

7.2 User Groups

Eye gaze as a communication device is more essential to some than others. There are those that can benefit from its use as their primary communication device and others to which it may improve or aid their communication. Additionally, there are those of which eye-gaze may be of little or some use now, but would become more in need of in the future.

Although, not intended as a scale, Hansen et al. (2006) used a similar approach in defining the accessibility requirements of ALS patients based upon the different stages of their condition. The first of five stages seeing symptoms such as a reduction in mobility with a suggestion that keyboards with added arm rests may be suitable. The fourth seeing symptoms of full-lock in with only eye movements suggesting that eye tracking may be suitable. The fifth seeing full-lock in with no eye movements suggesting that some form of bio-potential switch may be suitable.

Here, we are considering eye-gaze for communication as a sliding scale, see Figure 7.1, with us placing all of our experiment participants into one of three different user groups.

- **Group A:** This group is formed of individuals that would benefit greatly from eye-gaze and could be used as their primary means of communication. This group consists of young people, aged 8 to 17, suffering from some form of cerebral palsy (CP). Participants from this group are referred to as GP-AX ('X' being a participant number).
- **Group B:** This group is made of individuals who do not necessarily need to use eye-gaze as a primary communication device at this moment in time, but it may

aid their communication or benefit them in the future due to the degenerative nature of their condition. This group consists of young people, aged 15 to 18, suffering from muscular dystrophy (MD). Participants from this group are referred to as GP-BX ('X' being a participant number).

- Group C: This group is made made of individuals who do not need to use eye-gaze as a communication device. This group have no difficulties in communication and are able to use all conventional computer input devices. It consists of university students, aged 22 to 58, all are able-bodied with no physical disability. Participants from this group are referred to as GP-CX ('X' being a participant number).

User groups A and B were made of students attending Ash Field School in Leicester, UK. The individuals were selected by the deputy head teacher, a senior support teacher and the assistive technology assessor. They have known all of the students participants throughout their schooling to date, and felt that they would benefit in some form whether it be for fun, motivational or educational reasons. All ethical issues concerning the experimental procedures were addressed by the school, and all students and parents were consulted prior to the study.

Ash Field School is a day and weekly boarding special school for pupils of all abilities from five Local Authorities whose main presenting difficulties are physical. There are currently 112 pupils aged from 4 to 19 that are supported by 99 staff members. All pupils have a physical disability, a range of learning needs and a significant number have progressive conditions. A third of the pupils do not use spoken language as their primary means of communication.

Participants in Group A

Group A consists of 2 females and 2 males aged between 8 and 17; all are severely physically disabled. They all have little or no verbal communication and require some form of assistive device in order to communicate or use a computer. Due to the low number of people in this group, experimental design is based on case study. During the past year, they have all spent short periods of time using a Dynavox EyeMax eye tracker.



Figure 7.2: GP-A1 Walking around in World of Warcraft. Due to this person's condition, he naturally falls into the divergent sideways positioning as shown in the picture. His head restraint often covers one of his eyes from view of the eye tracker.

GP-A1 is male, 13 years old and suffers from athetoid cerebral palsy. His condition is such that he is always sitting in a divergent position, see Figure 7.2. Most of his body movements are the result of involuntary actions and any attempts at deliberate movements are erratic. To ease his dystonia (uncontrollable and painful muscle contractions) symptoms, he was implanted with a deep brain stimulator (DBS). He is unable to speak but able to vocalise and use facial expression for yes and no answers. His primary communication is through an Etran frame (Figure 2.8) and a Dynavox electronic communication aid (Figure 2.9) operated using a switch that is activated using his left foot.

GP-A2 is female, 14 years old and suffers from Glutaric Aciduria Type 1 (GAT1). This disorder affects the body's ability to completely breakdown certain amino acids. The partially broken down product accumulates and causes damage to areas of the brain including the basal ganglia, which is responsible for regulating movement. In order to help with posture and to prevent her spine from curving she has had spinal rods surgically implanted. Most of her body movements are the result of involuntary actions and any attempts at deliberate movements are erratic. She has difficulty in speaking but is able to communicate slowly with single words; this requires much ef-

fort. Her communication is supported by the use of a Dynavox communication aid which is operated using a single head switch mounted to her wheelchair. She is able to use the vocabulary appropriately for formal and informal communication and actively takes part in class discussions. She is very independent but not confident in using new technology as she is easily frustrated which, usually results in her becoming demotivated.

GP-A3 is female, 8 years old and suffers from cerebral palsy. She has minimal muscle tone and struggles with all types of body movement. When at full health she is able to use a touch pad switch but is quickly fatigued after several minutes. She is unable to speak but is able to vocalise and smile when she is motivated. She is able to use an Etran frame for class activities and is able to make clear choices through eye pointing and to answer questions.

GP-A4 is male, 17 years old and suffers from cerebral palsy. Most of his body movements are the result of involuntary actions and any attempts at deliberate movements are erratic. He has virtually no spoken language but can make vocal noises to give yes and no answers. He eye points effectively using an Etran frame, choosing from up to 10 symbols. He uses two switches operated by his hands to access a Dynavox communication device.

A summary of the individuals in Group A can be seen in Table 7.1.

<i>User</i>	<i>Age</i>	<i>Physical Conditions</i>	<i>Main Method of Communication</i>	<i>Input Devices Used</i>
GP-A1	13	CP	Etran Board, Dynavox communicator	Foot switch
GP-A2	14	GAT1	Etran Board, Dynavox communicator	Head switch
GP-A3	8	CP	Etran Board, Dynavox communicator	Touch pad switch
GP-A4	17	CP	Etran Board, Dynavox communicator	Head switch

Table 7.1: Summary of the participants in Group A

Participants in Group B

Group B consists of 8 males aged between 15 and 18; all suffering from muscular dystrophy. Their current condition means that they all require a wheel chair for mobility that is controlled via a joystick. They can all communicate effectively verbally and are able to use a mouse and keyboard however, it is at a slow pace with poor accuracy and often for only short periods of time. Low screen resolutions are typically used to allow for inaccuracies when using the mouse as many of the individuals are restricted in some of their hand and arm movement. Some individuals will use keyboards with large keys or have key guards fitted to prevent accidental key strokes. None of the individuals use a communication device in their classes and none have previously used an eye tracker.

Participants in Group C

Group C consists of 15 males and 5 females, able-bodied mouse and keyboard users. All were either students or staff members from De Montfort university.

7.3 Experiment 1: Understanding User Requirements

This first series of case studies is to establish the capabilities of the CP participants in Group A. The main research questions being:

- Can the participants in Group A perform the gaze interaction techniques that have been developed and proven by able-bodied users to allow control in MMO environments (see Chapter 5 and Chapter 6)?
- Does their performance differ and if so by how much?
- What difficulties (if any) do the participants experience?
- What are the range of tasks that CP participants in Groups A are able to perform?

7.3.1 Configuration

World of Warcraft was used as the virtual environment. World of Warcraft and Snap Clutch were run from a Dell Latitude D830 laptop with 4GB of RAM and 1GB of Video RAM. Output from the laptop was sent to a 20 inch wide screen Samsung 206BW monitor set at a resolution of 1680x1050 px with a Tobii X120 eye tracker positioned

7.3 Experiment 1: Understanding User Requirements

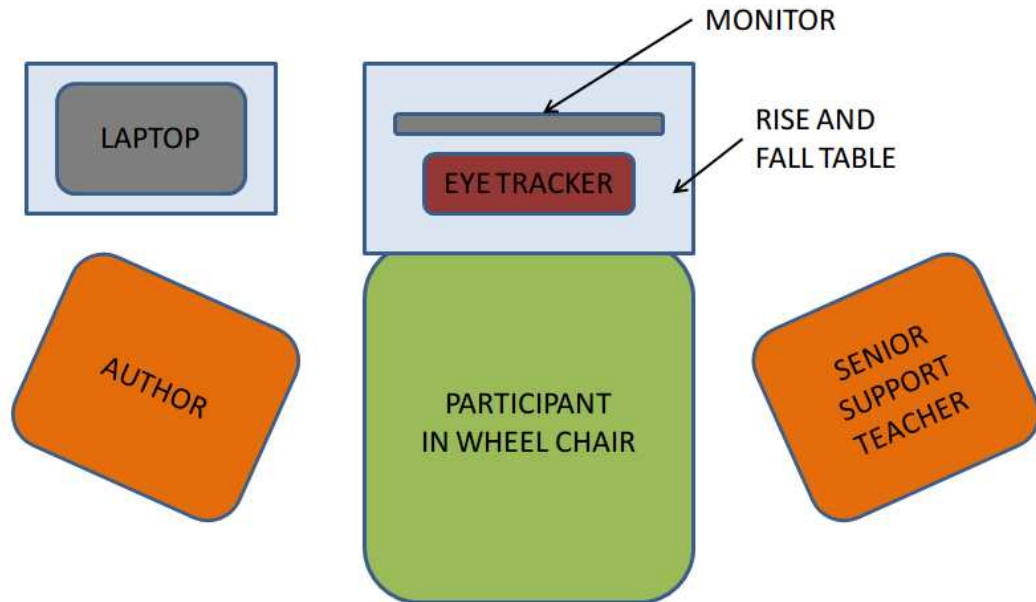


Figure 7.3: Plan view of the experiment setup used

below. The monitor and eye tracker were placed on a rise-and-fall table. The table height and eye tracker positioning angle was adjusted to suit each participant, Figure 7.3 shows the experiment arrangement.

Two different Snap Clutch configurations were to be used. The first configuration included a set of modes that allowed the individual to control locomotion only:

- Glance Up: 'Locomotion' mode
- Glance Down: 'Look around' mode
- Glance Left: No action
- Glance Right: No action

The 'no action' mode was applied to the left and right glances, so as to limit the participants to the two modes that are applied to the top and bottom glances. The second configuration included a set of modes that allowed the individual to control locomotion and interact with the user interface:

- Glance Up: 'Locomotion' mode

7.3 Experiment 1: Understanding User Requirements

- Glance Down: ‘Look around’ mode
- Glance Left: ‘Left mouse dwell click’ mode
- Glance Right: ‘Right mouse dwell click’ mode

7.3.2 Procedure

This study was carried out with participants in Group A over two sessions at Ash Field School, several weeks apart. To assist with the study a senior support teacher familiar with the children taking part was present throughout. Each individual was given a simple explanation of what the eye tracker is and how it was going to allow them to play a computer game by only looking around the screen. It was presented as informally as possible so as to not apply any kind of stress. After eye tracker calibration each participant was free to choose and design their own character to represent themselves in World of Warcraft. This was done by the experimenter using the mouse and keyboard under yes/no instruction from the individual. Following, Snap Clutch was started within the game and it was explained how the interaction technique for walking about in the world works. The participant was then free to explore the environment as they wished. Sessions were not video taped, they were recorded using observation sheets developed by the school and included notes on:

- Participant
- Session
- Session length
- Current physical and emotional state
- Observations
 - Positioning
 - Calibration
 - Task performance
 - Response

Following, is a descriptive summary and discussion of the observations made.

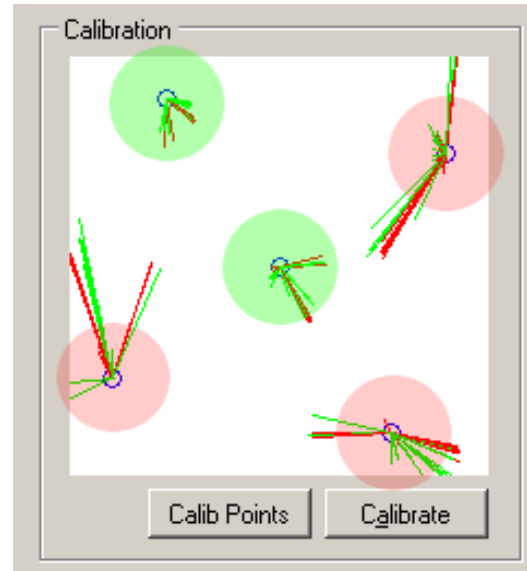


Figure 7.4: Experiment 1: Calibration quality was assessed visually. This Figure shows an example of a poor calibration with only two out of five points visually acceptable. Individual points are recalibrated to improve the overall calibration.

7.3.3 Observations

GP-A1

The individual arrived at both sessions in positive emotional health but not physically. That is, he was excited and motivated to take part but would quickly fatigue. Due to his DBS only being implanted recently, he must undergo a large amount of physiotherapy which causes him to tire quickly. Calibration and eye tracker positioning for both sessions was difficult due to his divergent positioning but was achieved after approximately 10 minutes. Acceptable/poor calibrations was determined visually using the Tobii calibration map, see Figure 7.4. Only one eye was tracked. Once in the game the participant had difficulties in starting and stopping character movement. This was due to the eye tracker losing the eye it was tracking due to involuntary head movements. Additionally, there were occasions when the head movement would result in both eyes being visible to the eye tracker for very short periods of time and the software then trying to estimate new gaze positions. This caused additional latency. On both sessions, this individual became tired after approximately 5 minutes of gameplay.

7.3 Experiment 1: Understanding User Requirements

GP-A2

Prior to the sessions, the individual had been responding well during class and was feeling physically well. She was feeling a little anxious for the first session but was excited for the second. A successful calibration was achieved on the third attempt for the first session and second attempt for the second session, this was due to involuntary head movement and moving beyond the head box of the eye tracker. Issues with head movement would continue upon entering the world with which she struggled to move her character using the locomotion technique. She was able to start and stop character movement and also walk around but without any real control. After several adjustments to the individuals wheel chair head support, it was necessary for us to offer hand-head¹ support. After head support assistance, she was able to move around the world as she pleased. As she relaxed during the session, her head was slowly released until she was able to support herself using the wheel chair head support. During the session her involuntary head movement seemed to also reduce. For the second session she was given navigational instructions for exploring nearby points-of-interest: a farm; an old fort and also a small camp site. She was able to navigate to these areas but not directly and would often be caught in obstacles such as trees and would walk too far or turn too much. Each session after calibration lasted around ten minutes.

GP-A3

The individual arrived at both sessions in good physical and emotional health. She appeared very relaxed but not particularly anxious or excited for the first session but very excited for the second session. As this individual often leans her head back at almost 45 degrees, it was necessary to adjust her head support to raise her head enough so that the eye tracker could clearly see her eyes. For the first session hand-head support was provided but this was not required for the second session. An acceptable calibration was achieved for both sessions on the second attempt. The character creation process was a very emotive and enjoyable one for this individual. Upon scrolling through different character appearances she would verbalise on ones that she wanted to select. When in the game, she was able to start and stop character movement but not able to

¹Hand-head support means to use our hands to hold the individual's head in place to prevent it from falling forwards.

7.3 Experiment 1: Understanding User Requirements

turn left and right. When realising that she was making her character start and stop walking she began to vocalise loudly and opened her eyes widely. To accommodate for any lack of understanding (this individual is only 8 years old), the sides of the screen were pointed to in order to try and direct her gaze so as to rotate her character. However, she suffered from a lot of involuntary head movement causing the eye tracker to lose her eyes. For the second session, her mother attended and she commented on the high amount of positive verbal expression made by the individual.

GP-A4

Prior to both sessions the individual was in good physical and emotional health. He was very excited to be taking part in the study and also playing a computer game such as World of Warcraft for the first time. During both sessions, the participant found the calibration process to be quite stressful, with the result often being not enough data collected for each calibration point. Thus, it was necessary to calibrate one of the five points at a time until enough calibration data was collected to continue. During this time the individual would suffer from a large amount of head movement, arm movement and was also quite verbal. Before each calibration point, it was necessary to check if he was happy to continue or wished to stop. Once the game was started the participant relaxed and made much less head and arm movement. He was able to start and stop his character but was only able to turn left during both sessions. Occasionally, it was necessary to provide additional hand-head support as his head would often fall forwards. Towards the end of the second session he walked into a river in game and stopped his character leaving it floating in the water. At this point, the individual appeared to completely relax his body in his chair and made virtually no body movements.

Observation Conclusions

This first set of observations immediately demonstrates that locomotion in realistic environments is a motivating and possibly overwhelming experience for this group of participants. All participants found both sessions engaging, enjoyable and motivating. However, the verbal and physical communication received from individuals indicates that the calibration process is stressful however, playing the game is a much more relaxing experience. This suggests that the environment in which the game is set (land, water, forest etc) and the state of the character (walking, running, swimming

7.4 Determining Performance of a Gaze Interaction Technique

etc) has an effect on the individual's ability to relax. The difficulty in setup, calibration and character control does demonstrate that there is a need for dynamic accessibility that considers a person's current health and condition. Anxiety and stress has an effect on character control and GP-A2 showed that as her confidence increased, her anxiety reduced, lowering her stress level resulting in more successful character control. This resulted in greater relaxation and subsequently a reduction in involuntary head movement. The success greatly improved her motivation for the second session and she was able to follow navigational instruction. An increase in motivation and engagement was also observed with GP-A3 whom is normally quiet and non-verbal when in class, sometimes appearing to lack motivation. However, during both sessions it was observed by ourselves and her mother that the experience was highly engaging and motivating despite her having little control over her character.

These observations are summarised in Table 7.2 with definitions of the attributes used in Table 7.3. The performance column is split into three different tasks and is intended to be a simple evaluation on the user's ability to control their character. The tasks were:

1. Can start/stop their character moving? (Yes/No)
2. Can follow instructions? (Yes/No)
3. Can move their character in all directions as required? (Yes/No)

The expectation was that CP participants in Group A would be able to perform character locomotion using the interaction technique, but this was observed not to be the case. The next step is to consider the observations and to try and determine a way of measuring them such that the interaction technique can be modified to better suit them.

7.4 Determining Performance of a Gaze Interaction Technique

The first series of case study observations in experiment 1 presented two key questions:

- How can we measure a user's ability to perform a particular interaction technique?
- Can we adapt the interaction technique to better suit the individual?

		Prior State		Positioning			Sess. Length		Performance		
<i>Participant</i>	<i>Sess- ion</i>	<i>Emot- ional</i>	<i>Physi- cal</i>	<i>Head Support</i>	<i>Head Control</i>	<i>Calib. Quality</i>	<i>Setup</i>	<i>Game Play</i>	<i>Task 1</i>	<i>Task 2</i>	<i>Task 3</i>
GP-A1	1	Excited	Poor	Poor	Good	Good	10m	5m	Y	N	N
GP-A1	2	Excited	Poor	Poor	Good	Poor	10m	5m	N	N	N
GP-A2	1	Anxious	Good	Medium	Good	Good	10m	10m	Y	Y	Y
GP-A2	2	Excited	Good	Medium	Good	Good	5m	10m	Y	Y	Y
GP-A3	1	Relaxed	Good	Medium	Poor	Good	10m	10m	Y	N	N
GP-A3	2	Excited	Good	Medium	Good	Good	10m	10m	Y	N	N
GP-A4	1	Excited	Good	Poor	Poor	Good	15m	10m	Y	N	N
GP-A4	2	Excited	Good	Medium	Poor	Good	5m	10m	Y	N	N

Table 7.2: Summary of observations made during Experiment 1. See Table 7.3 for definitions and level descriptions of attributes.

7.4 Determining Performance of a Gaze Interaction Technique

<i>Attribute</i>	<i>Level</i>	<i>Definition</i>
Head Support	Poor	Needs extra support
	Medium	Uses head restraint
	Good	No support needed
Head Control	Poor	Eye tracker loses eyes often
	Good	Eyes visible to eye tracker most of the time
Calibration Quality	Poor	< 3 calibration points (see Figure 7.4)
	Good	>= 3 calibration points (see Figure 7.4)
Performance	Task 1	Can start and stop their character moving? (Yes/No)
	Task 2	Can follow instructions? (Yes/No)
	Task 3	Can move their character in all directions as required? (Yes/No)

Table 7.3: Definition of attributes and levels found in Table 7.2

<i>Hardware</i>	<i>Environment</i>	<i>User</i>
Eye tracker accuracy	Lighting	Head control
Eye tracker tolerance to head movement		Head positioning
Eye tracker latency		Eye construction
Embedded software		Eye control
		Understanding

Table 7.4: Influences on gaze interaction performance

7.4.1 Influencing Factors

The results shown in Table 7.2 are subjective and do not offer sufficient information that could be used to improve the individuals performance. However, they can be used to define the factors that influence performance, and are categorised as in Table 7.4. Firstly, there are hardware and embedded software (such as glint detection, calibration algorithms and so on) constraints with the eye trackers ability in estimating the users gaze point; secondly, there are user related factors and thirdly, environmental factors.

The poorer these attributes are then the noisier the data with which, the interaction

7.4 Determining Performance of a Gaze Interaction Technique

technique must work with. Although, we have little or no control over these attributes we can design variations of the interaction technique that allow them to be more tolerant to poorer attributes, see Section 5.3 for a discussion on possible methods of varying interaction techniques.

One method to quantify ability and ultimately improve performance is by developing a metric and diagnostic test that measures the influencing factors. The diagnostic test is based upon the interaction technique and the requirements for completion should be equal to what is required to operate the interaction technique. Considering dynamic accessibility, the test should measure an individuals ability to complete the test when using a specific eye tracker, albeit of low or high quality (quality being defined by the hardware influencing factors) at a specific point in time. A higher quality eye tracker will be more tolerant to many of the user influencing factors than a lower quality eye tracker; additionally, an individual may perform differently depending upon the time of day, level of health, fatigue and so on. The metric used to determine the test results is thus based upon data collected from the eye tracker and the user's ability to complete the test rather than the individual factors and their relationships. This is considered to be a measure of a user's reliability (R_{user}) in performing the test rather than their success in completion. The data to be collected and the directly affecting factors are as follows:

- Percentage of time at least one eye is being tracked during the diagnostic test
 - Influenced by: eye tracker tolerance to head movement, eye tracker latency, head control, head positioning, eye construction, lighting
- The number of fixations made to complete the diagnostic test
 - Influenced by: eye tracker accuracy, eye tracker latency, head control, eye construction, eye control, understanding
- The time taken to complete the diagnostic test
 - Influenced by: eye tracker latency, eye control, understanding

7.4.2 A Diagnostic Test Based Upon the Interaction Technique Requirements

As stated previously, the focus of this series of experiments is that of locomotion and so the interaction technique that we want to assess an individual's reliability in performing is the transparent overlay (see Subsection 5.2.1). The interaction technique's requirement of the user is that they must be able to perform at least a single fixation within all of the interaction elements found on the screen (see Table 5.1). Therefore, the diagnostic test is required to measure that individuals ability to fulfil these requirements. The test is constructed as follows:

- The individual is presented with a blank, black screen and a countdown timer to indicate the start of the diagnostic test
- Once the timer has complete an on-screen button will appear within a random time interval between 0-5 seconds
- The on-screen button will be in one of the positions and of the same size as one of the interaction elements (defined by the interaction technique)
- The individual must then perform a single fixation (defined by the interaction technique) within the on-screen button
- If the user is fixating when the test sequence has started then this fixation is ignored
- If successful the on-screen button turns green and disappears; if a time-out is reached then the on-screen button disappears
- A new on-screen button will then appear within a random time interval between 0-5 seconds
- There are nine possible on-screen button positions (defined by the interaction technique) and each position appears twice during the test; thus a total of 18 test sequences
- The order that the on-screen buttons are presented to the individual are random

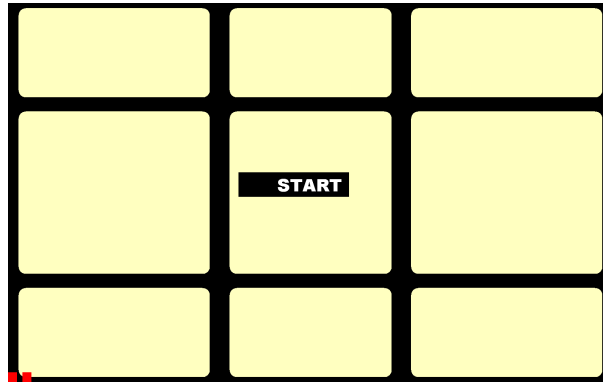


Figure 7.5: Screenshot of the diagnostic test used to assess user reliability for the locomotion interaction technique. The position and size of the buttons was defined by the interaction technique used for locomotion.

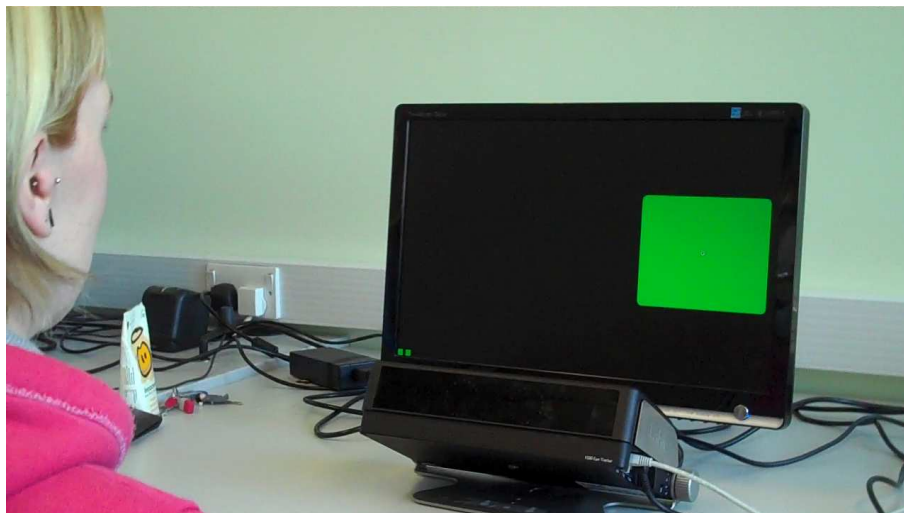


Figure 7.6: Diagnostic test in progress.

A screenshot showing the arrangement of the on-screen buttons can be seen in Figure 7.5 and an image taken of a diagnostic test in progress can be seen in Figure 7.6

During the test all eye-gaze data is recorded for calculating reliability and for re-playing gaze scan paths.

7.4.3 Developing a Metric of Reliability (R_{user})

Reliability is based upon the following:

‘The longer the time taken to complete the test sequence then the poorer the reliability’

‘The more fixations made by an individual when performing a test sequence then the longer the time taken to complete the test sequence; and subsequently the poorer the reliability’

‘The longer the length of time that the eyes are not being tracked then the longer the time taken to complete the test sequence; and subsequently the poorer the reliability’

One important consideration is for the number of fixations made and the percentage of time the eyes are being tracked. Here, we have a case of one not existing with the other, for example, if the eyes are not being tracked then it is impossible to measure and determine the number of fixations made; likewise, if fixations are being measured then the eyes *are* being tracked. Thus, only one of these data measures can be used within the metric at any one time. Considering that these are mutually exclusive and we are determining overall reliability, the choice of which data measure to use must default to the lower of the two.

Thus, the user’ metric of reliability is defined as follows:

$$R_{user} = \begin{cases} \frac{R_{time} + R_{fixation}}{R_{time} + R_{eyes}} & \text{if } R_{fixation} < R_{eyes} \\ \frac{R_{time} + R_{eyes}}{R_{time} + R_{fixation}} & \text{if } R_{eyes} < R_{fixation} \end{cases} \quad (7.1)$$

Where R_{time} is a measure of reliability relative to the time taken to complete the test sequence; $R_{fixation}$ is a measure of reliability relative to the number of fixations made; and R_{eyes} is the percentage of time that the eyes were visible by the eye tracker during the test sequence.

7.4 Determining Performance of a Gaze Interaction Technique

Example ($R_{time} = 77.8\%$, $R_{fixation} = 28.6\%$, $R_{eyes} = 83\%$):

$$R_{fixation} < R_{eyes} \therefore R_{user} = \frac{77.8\% + 28.6\%}{2} = 53.2\% \quad (7.2)$$

Re-factoring as a percentage and making the assumption of linear progression (see Figure 7.8), R_{time} is defined as:

$$R_{time} = \frac{T_{max} - T_{user}}{\frac{T_{max} - T_{min}}{100}} \quad (7.3)$$

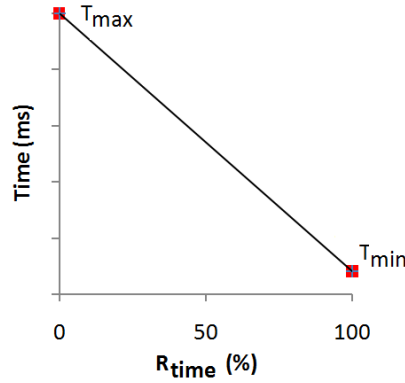


Figure 7.7: Re-factoring time as a percentage to calculate R_{time} , see equation 7.3

Where T_{max} is the maximum time allowed to perform a test sequence; T_{min} is the minimum time required to perform a test sequence; and T_{user} is the individuals actual time when performing the test sequence.

Example ($T_{max} = 10000$, $T_{min} = 250$, $T_{user} = 2416$):

$$R_{time} = \frac{10000 - 2416}{\frac{10000 - 250}{100}} = 77.8\% \quad (7.4)$$

Re-factoring as a percentage and making the assumption of linear progression, $R_{fixation}$ is defined as:

$$R_{fixation} = \frac{F_{max} - F_{user}}{\frac{F_{max} - F_{min}}{100}} \quad (7.5)$$

7.4 Determining Performance of a Gaze Interaction Technique

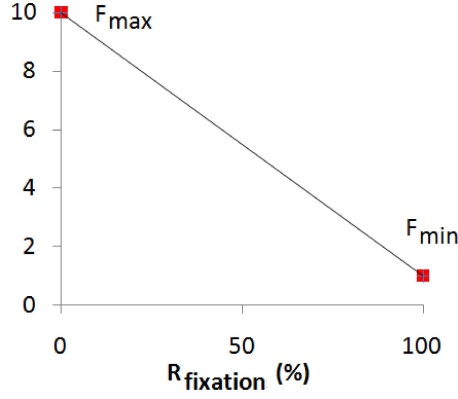


Figure 7.8: Re-factoring time as a percentage to calculate $R_{fixation}$, see equation 7.5

Where F_{max} is the maximum allowable fixations required to perform a test sequence; F_{min} is the minimum number of fixations to perform a test sequence; and F_{user} is the individuals actual number of fixations made when performing the test sequence.

Example ($F_{max} = 12$, $F_{min} = 5$, $F_{user} = 11$):

$$R_{fixation} = \frac{12 - 11}{\frac{12 - 5}{100}} = 28.6\% \quad (7.6)$$

R_{eyes} is defined as:

$$R_{eyes} = \frac{T_{eyes}}{T_{user}} \quad (7.7)$$

Where T_{eyes} is the length of time that the eyes were visible by the eye tracker during the test sequence and T_{user} is the length of time to complete the test sequence.

Example ($T_{eyes} = 3360$, $T_{user} = 4032$):

$$R_{eyes} = \frac{3360}{4032} = 83\% \quad (7.8)$$

Setting Parameters to Determine T_{max} , T_{min} , F_{max} , F_{min}

T_{min} and F_{min} can be set by the interaction technique requirement; that is, the minimum time to initiate an interaction element is a single fixation of 80 ms. In order to determine the values for T_{max} , F_{max} we performed an initial evaluation of the diagnostic test. The test configuration is described in detail in Section 7.5.

7.4 Determining Performance of a Gaze Interaction Technique

Five able-bodied participants were recruited from within the Faculty to perform the test, all had taken part in previous eye-gaze studies but were all novice users of the locomotion interaction technique. The test procedure was explained to them prior to the start of the test. All five completed the diagnostic test with no difficulties. The results are as follows. The \overline{F}_{user} was 2.6 fixations and the \overline{T}_{user} was 496 ms. By rounding the Figures and considering a factor of 5 for the maximum number of allowable fixations and a factor of 10 for the maximum allowable time we can summarise all parameter values as follows:

- $T_{min} = 80$ ms
- $T_{max} = 5000$ ms
- $F_{min} = 1$
- $F_{max} = 10$

Simplifying the Metric

Each factor within the reliability metric is given an equal weighting. This is intended to be a starting point and not definitive as there will most certainly be a relationship of different weights between each one of the factors. Establishing the relationships will increase the accuracy and robustness of the metric. Although this goes beyond the scope of this thesis it is considered an important piece of future work that can be achieved by isolating each of the factors as independent variables.

7.4.4 Adapting the Interaction Technique

The diagnostic test described previously results in a reliability measure that determines the overall reliability of an individual performing the transparent overlay interaction technique using a particular quality of eye tracker. It also provides the reliability for the individual interaction elements and their positions on screen. Using this information it is possible to move the interaction elements and their associated functions to areas of the screen that the individual has a higher reliability. Of course, moving the interaction elements is not the only way to adapt the interaction technique, see Section 5.3, and the adaption could go further by changing the size of the interaction elements. However, as a first step, we want to verify that moving the elements has a positive affect on

7.4 Determining Performance of a Gaze Interaction Technique



Figure 7.9: The priority levels of character control. The first objective of any user is to get their character moving; Level 1. Those that can fully use the locomotion interaction technique are capable of Level 6 control, that is full control.

character control. The virtual world task that we want users to accomplish is that of controlled locomotion. This requires the user to navigate their way in the world and move their character forwards, left and right. By applying a priority system to the control functions of the interaction elements we can determine their new placements based upon the individuals reliability. The priority control system is based upon the different levels of character control and their significance, see Figure 7.9.

The first priority is to get the character moving; Level 1. This then increases by adding more essential controls until all of the controls requirements are met; Level 6. From the observations in experiment 1 we have one user capable of ‘Level 6’ control; two users that are at ‘Level 2’ control and one at ‘Level 1’. The scope between ‘Level 5’ and ‘Level 6’ controls is minimal as there is justification to eliminate the need for reverse movement altogether. The low level event data captured during evaluation study 3A (see Subsection 6.2.2) indicated that there were no recorded instances of backward movement. However, it is a part of the available control system and so is included in the levels of control.

The priority function system can be seen in Figure 7.10 and Table 7.5; it operates as follows: Users that achieve a $\overline{R}_{user} \geq 70\%$ should be capable of using the transparent overlay interaction technique without any modification. A threshold is certainly required but defining that value accurately is not possible at this stage. The value cannot be too high so as not to be reached nor too low that is easily achieved. Therefore, a middle-to-top value of 70% is used as a starting point. If a user achieves a $\overline{R}_{user} < 70\%$ then we should proceed and modify the technique. With reference to Figure 7.11 and taking an extreme worked example, if we have a disabled participant who can only move their eyes vertically (e.g. a stroke patient suffering from gaze paresis/palsy, see

7.4 Determining Performance of a Gaze Interaction Technique

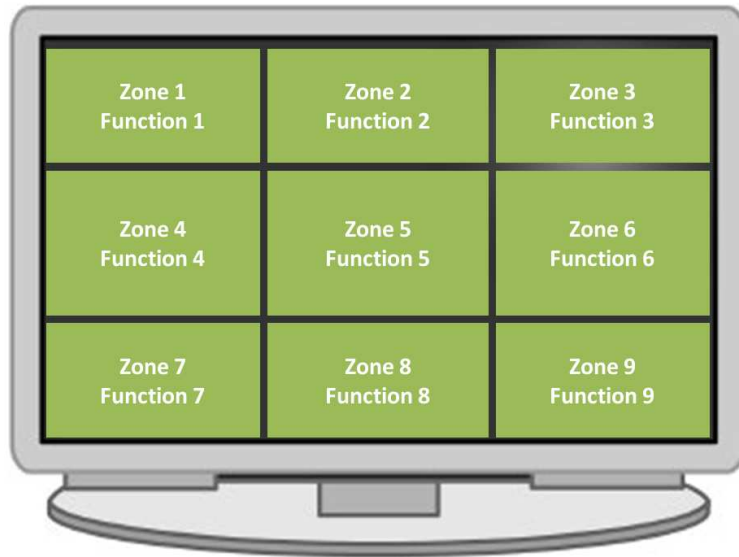


Figure 7.10: Zone and function positions for the transparent overlay interaction technique

<i>Priority</i>	<i>Control</i>	Function Zone Position Preference			
		<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>
1	Forward	5	8	2	4 or 6
2	Left	7	4	1	8 or 2
2	Right	9	6	3	8 or 2
3	Forward/left	4	1	x	x
3	Forward/right	6	3	x	x
4	Backwards	8	x	x	x

Table 7.5: Overview of the priority system to which functions are assigned to zones

7.4 Determining Performance of a Gaze Interaction Technique

Subsection 2.4) then the expectation might be for them to score R_{user} values $> 70\%$ for zones 2, 5 and 8 and score $< 70\%$ for zones 1, 3, 4, 6, 7 and 9. A partial step through the the priority function system for this participant can be seen in Table 7.6 and the adapted interaction technique in Figure 7.12. The result being that this individual can now use an adapted version of the interaction technique that allows him to perform basic locomotion to ‘Level 3’.

<i>No.</i>	<i>Flow No.</i>	<i>Instruction</i>	<i>Outcome</i>
1	1	Go to first priority control	Assigning Forward control
2	2	Is preferred zone available?	Yes, preferred zone is 5
3	3	Assign zone to control	Forward control assigned to Zone 5
4	4	Any more zones to assign?	Yes
5	5	Is there another priority control next?	Yes, Left control is next
6	2	Is preferred zone available?	No, preferred zone is 7
7	6	Is there a next preferred zone available?	Yes
8	2	Is preferred zone available?	No, preferred zone is 4
9	6	Is there a next preferred zone available?	Yes
10	2	Is preferred zone available?	No, preferred zone is 1
11	6	Is there a next preferred zone available?	Yes
12	2	Is preferred zone available?	Yes, preferred zone is 8 or 2
13	3	Assign zone to control	Left control assigned to Zone 2
14	4	Any more zones to assign?	Yes
...

Table 7.6: An example of how the priority Zone/Control selection system operates in reference to the process flow chart in figure 7.11. Note, that this is only a partial walk through the process and thus only two Controls have been assigned. The adapted interaction technique appears as in Figure 7.12.

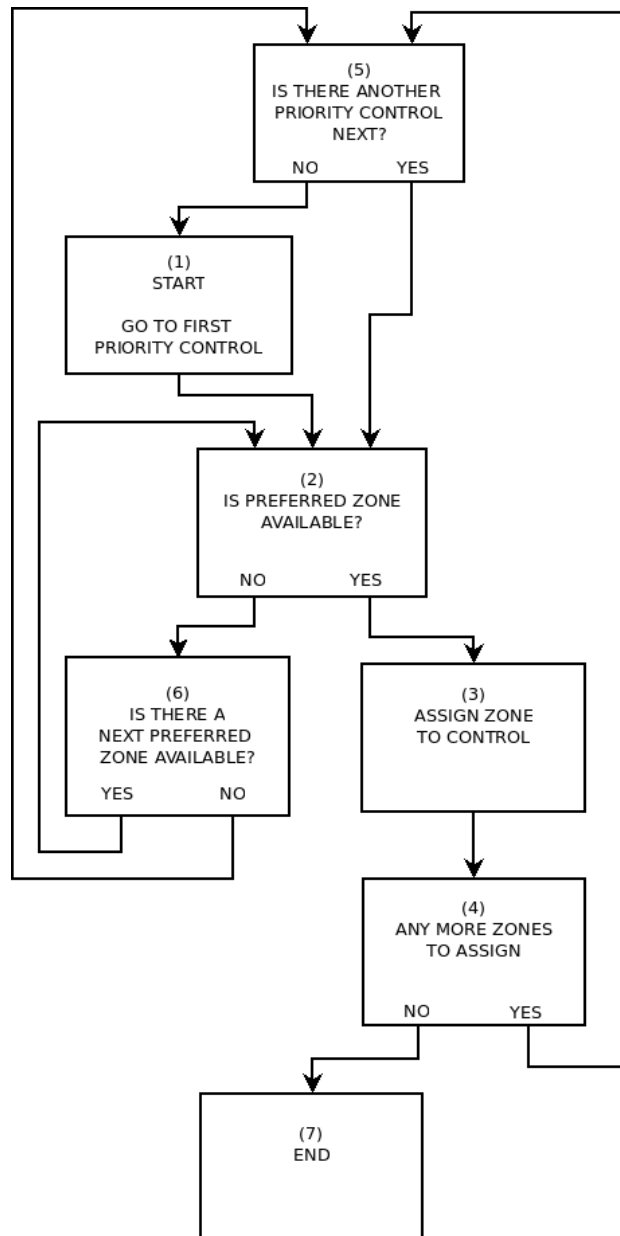


Figure 7.11: The process flow for allocating controls to zones based on the priority system in Figure 7.10 and Table 7.5

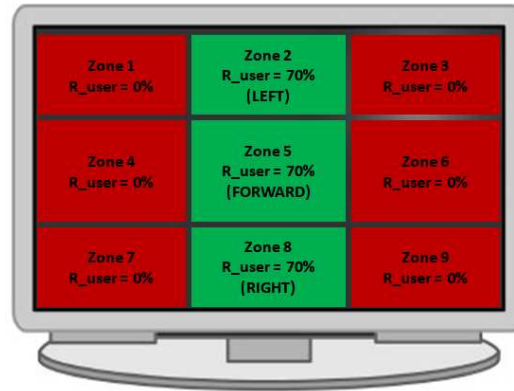


Figure 7.12: An example of an adapted interaction technique. Here, the participant is only able to move their eyes vertically and so scored 0% in Zones 1, 3, 4, 6, 7 and 9 and 70% in Zones 2, 5 and 8. By following the priority system (a partial step though can be seen in Table 7.6) three Controls are assigned as above.

7.4.5 Non-conventional Design

Gestalt¹ psychology suggests that the brain operates with a tendency to self-organise in a holistic, parallel way. The use of ‘gestalt laws’ (Sternberg, 2003) such as closure, similarity, proximity, symmetry and continuity are thus encouraged in user interface design. The unmodified interaction technique follows these laws in that the controls are presented symmetrical to one another. By applying the same laws to the modification of this interaction technique the amount of modification possible would be limited. Therefore, in this context the gestalt laws are considered lower in priority than the providing of function to the user. This may result in a user interface that appears unusual in appearance but is perfectly customised to suit the individual.

7.4.6 Fitts’ Law

This particular diagnostic test is similar in operation to a standard Fitts’ law target acquisition test. There have been several discussions on the application of Fitts’ law and if it applies to eye-gaze as an input device (e.g. Miniotas, 2000; Vertegaal, 2008; Zhang and MacKenzie, 2007). The main argument against it’s application is that eye movements are considered ballistic, that is, once started they cannot be stopped and the end point of a saccade is not known until it is reached. For Fitt’s law to apply both

¹Gestalt (German) - ‘an organized whole that is perceived as more than the sum of its parts’

7.5 Experiment 2: Diagnostic Test and Adaption of the Interaction Technique

the start point and end point need to be known. In the context of the diagnostic test presented here, the user does not have a start position. Further, Fitts' law is used to model the act of 'pointing' and in this application we want to model user reliability.

7.5 Experiment 2: Diagnostic Test and Adaption of the Interaction Technique

The purpose of experiment 1 was to determine R_{user} for each individual in Group A and if necessary adapt the transparent overlay interaction technique according to the priority system stated previously.

7.5.1 Participants and Configuration

The diagnostic test was written in C# and used the Tobii SDK and the GazeEngine.cs fixation detection engine, found within the Snap Clutch framework, see Section 5.1. The software was run from a Dell Latitude D830 laptop with 4GB of RAM and 1GB of Video RAM. Output from the laptop was sent to a 20 inch wide screen Samsung monitor set at a resolution of 1680x1050px with a Tobii X120 eye tracker positioned below. The monitor and eye tracker were placed on a rise-and-fall table. The table height and eye tracker positioning angle was adjusted to suit each participant, Figure 7.3 shows a typical arrangement. All four individuals in Group A were participants in this study.

7.5.2 Procedure

An explanation of how the tests worked was given and each individual was allowed to run through part of the test to facilitate understanding. Following, the test began as described in Section 7.4.2.

7.5.3 Analysis and Results

The results collected are summarised in Table 7.7 and in Figure 7.13 with supporting data shown in Appendix A.

In addition to $\overline{R_{user}}$ for each of the nine zones and an overall mean of the measure, the mean level of success is also shown for contrast. The number of correct answers

	$\overline{R_{user}}$ per Zone										
<i>Participant</i>	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	$\overline{R_{user}}$ (<i>total</i>)	$\overline{success}$ (<i>total</i>)
GP-A1	57%	77%	79%	84%	94%	64%	54%	9%	6%	58%	94%
GP-A2	54%	68%	52%	66%	85%	58%	84%	69%	88%	69%	94%
GP-A3	50%	29%	52%	61%	70%	51%	0%	0%	0%	35%	56%
GP-A4	70%	50%	3%	69%	63%	19%	4%	8%	19%	34%	39%

Table 7.7: Experiment 2 results with $\overline{R_{user}}$ for each zone, overall $\overline{R_{user}}$ and $\overline{success}$

7.5 Experiment 2: Diagnostic Test and Adaption of the Interaction Technique

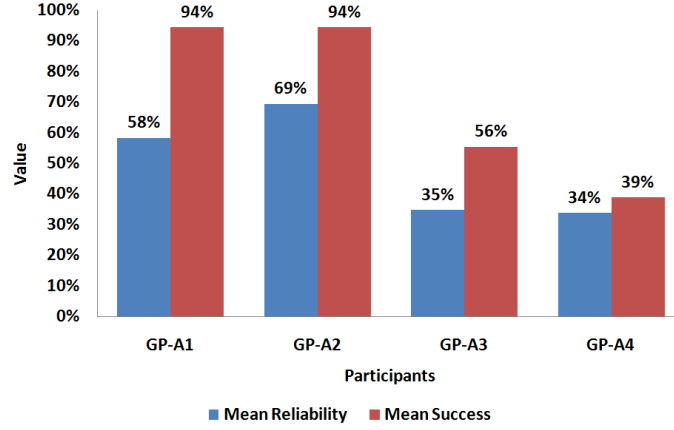


Figure 7.13: Experiment 2: Comparing ‘success’ as a measure with reliability

is often used as a measure of success in interactive tests such as this, but on it’s own does not offer any insight to how reliable the user can perform the tests; only that they were able to perform the test. For instance, both GP-A1 and GP-A2 were successfully able to fulfil the test sequence requirements in all but one instance giving them both a 94% success rate. However, their reliability in test sequence completion is reflected with their $\overline{R_{user}}$ value which, was much lower for both individuals. By creating a 3D surface map of the screen and applying the $\overline{R_{user}}$ values to the different zones we can visually see which parts of the screen a user can reliably perform this specific interaction technique. In the example found in figure 7.14, an image of a monitor has been overlay to indicate the orientation of the surface map.

Applying the priority function system, the surface maps can be used to visually replace the interaction elements of the interaction technique. Following this the controls for GP-A1 are assigned as Figure 7.15; for GP-A3 in Figure 7.16; and for GP-A4 in Figure 7.17. GP-A2 was able to control her character with the original interaction technique during experiment 1 and was only 1% away from 70% with her $\overline{R_{user}}$, therefore, it was considered not necessary to modify the technique. However, for reference, her surface map can be seen in Figure 7.18.

The surface maps produced indicate that there are certain areas of the screen that the participants had difficulty in accessing. More can be understood on the problems encountered by looking closer at the $R_{fixation}$ and R_{eyes} values as well as the gaze data collected.

7.5 Experiment 2: Diagnostic Test and Adaption of the Interaction Technique

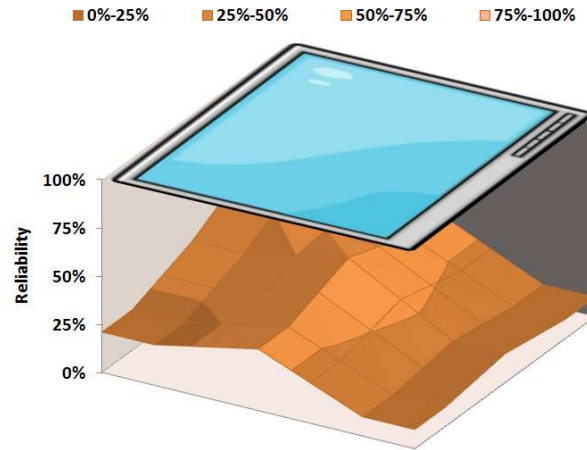


Figure 7.14: This shows an example surface map indicating $\overline{R_{user}}$ within the areas of the interaction technique elements. The overlaying image of the monitor indicates the orientation of the surface.

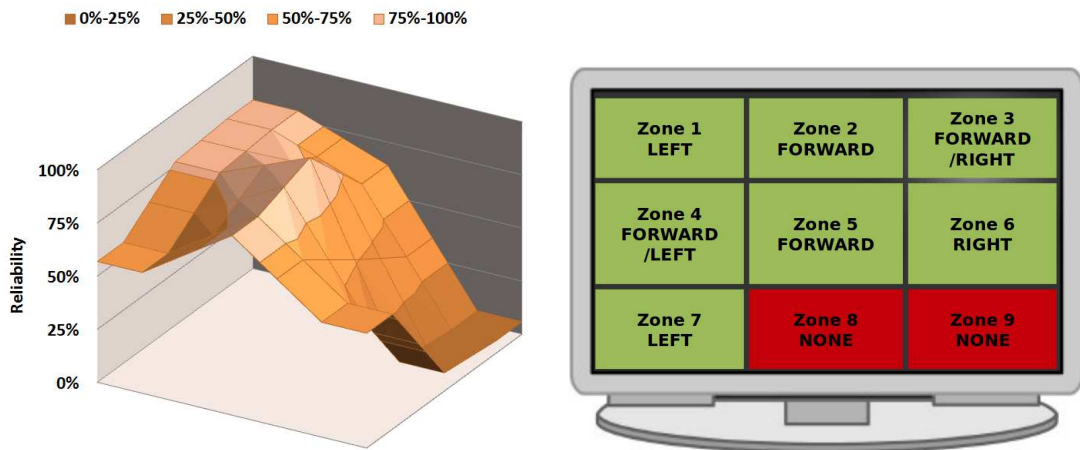


Figure 7.15: $\overline{R_{user}}$ surface map for GP-A1 (left) and their adapted interaction technique (right).

Prior to examining the failed scan paths, an example of a successful eye gaze scan path can be seen in Figure 7.19; both were performed by participant GP-A2 when attempting to acquire zones 2 (left) and 9 (right).

GP-A1, scored low $\overline{R_{user}}$ values for zone 8 (9%) and zone 9 (6%). The failure to complete the zone 8 test sequences can be explained to the individuals eyes being lost for 98% of the time, in contrast to zone 9 in which they were lost for only 28% of the time.

7.5 Experiment 2: Diagnostic Test and Adaption of the Interaction Technique

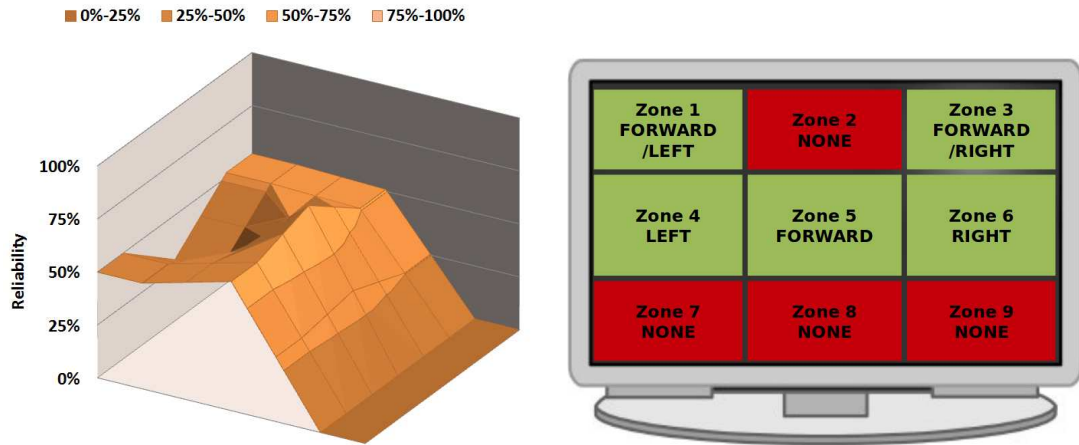


Figure 7.16: $\overline{R_{user}}$ surface map for GP-A3 (left) and their adapted interaction technique (right).

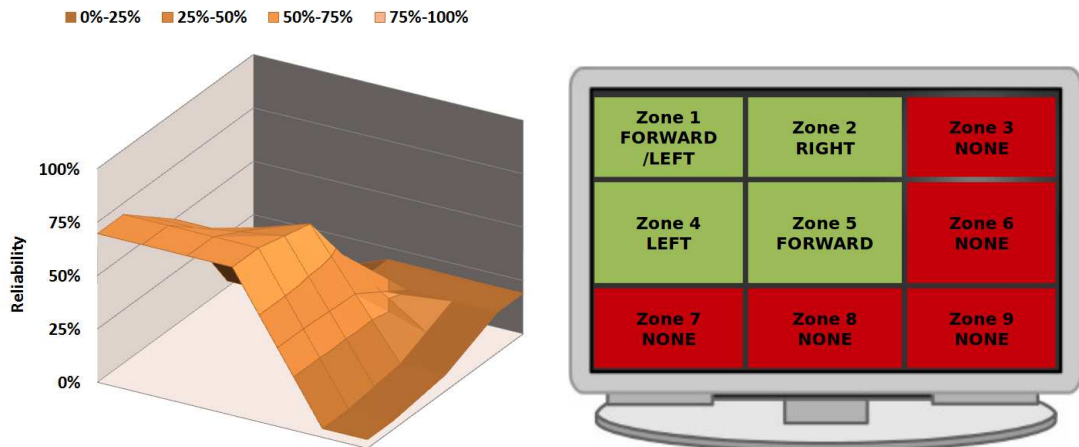


Figure 7.17: $\overline{R_{user}}$ surface map for GP-A4 (left) and their adapted interaction technique (right).

The large number of fixations made and the gaze scan paths, see Figure 7.20, suggest that the individual was attempting to fixate within zone 9 but the clustering indicates that the hardware influencing factors, such as eye tracker calibration or accuracy (see Table 7.4), for this region of the screen may be poor.

GP-A3, scored low $\overline{R_{user}}$ values for zone 2 (29%), 7 (0%), 8 (0%) and 9 (0%). The failure to complete zone 2 can be explained by the individuals eyes being lost for 79% of the time. However for zones 7, 8 and 9 the gaze scan path suggest that the user was

7.5 Experiment 2: Diagnostic Test and Adaption of the Interaction Technique

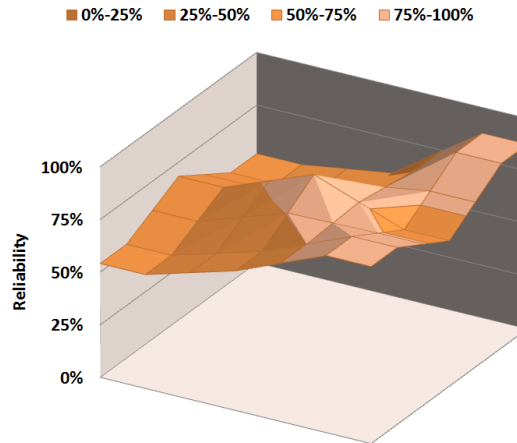


Figure 7.18: \overline{R}_{user} surface map for GP-A2

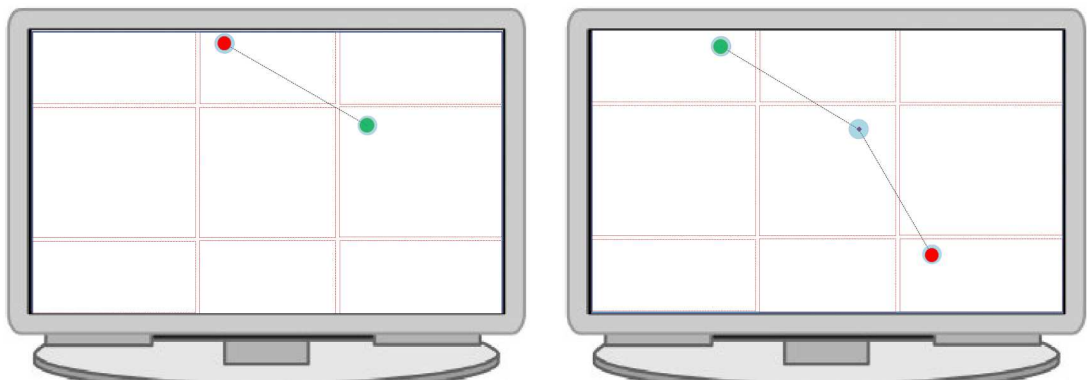


Figure 7.19: Successful attempts by GP-A2: Fixation eye gaze scan path for zone 2 (left) and for zone 9 (right). These two figures demonstrate the type of eye gaze scan path that is expected. The green circle indicates the first fixation and the red circle the last fixation made.

distracted and did not attempt to look within the correct zone. This was also noted during the study, where she would often look to me and watch what I was doing rather than watching the screen. This can be seen in Figure 7.21 (left) with which, prior to this sequence I attempted to bring her attention back by leaning across and touching the right side of the screen; however, she proceeded to follow my finger as I moved back to my position. This was repeated coincidentally throughout all sequences with zones 7, 8 and 9 and a similar scan path can be seen in Figure 7.21 (right) with which, the sequence started with her looking to me; then myself guiding her gaze back by touching

7.5 Experiment 2: Diagnostic Test and Adaption of the Interaction Technique

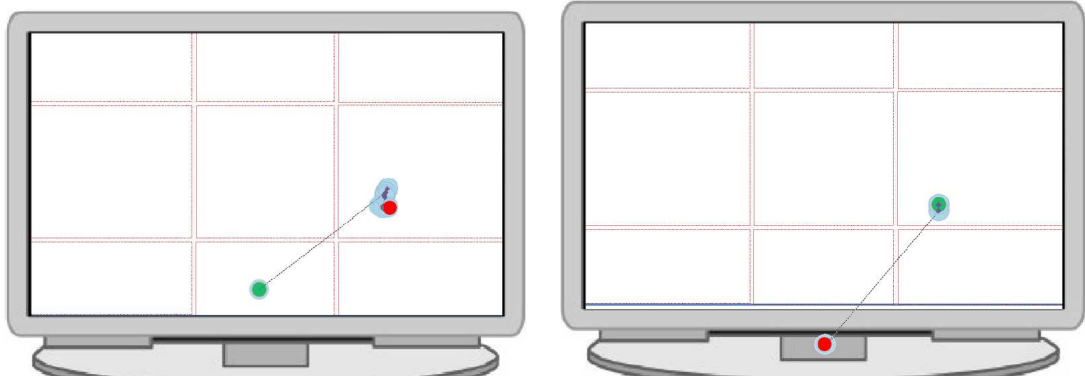


Figure 7.20: GP-A1: Fixation eye gaze scan path for zone 9 attempt 1 (left) and attempt 2 (right). The green circle indicates the first fixation and the red circle the last fixation made.

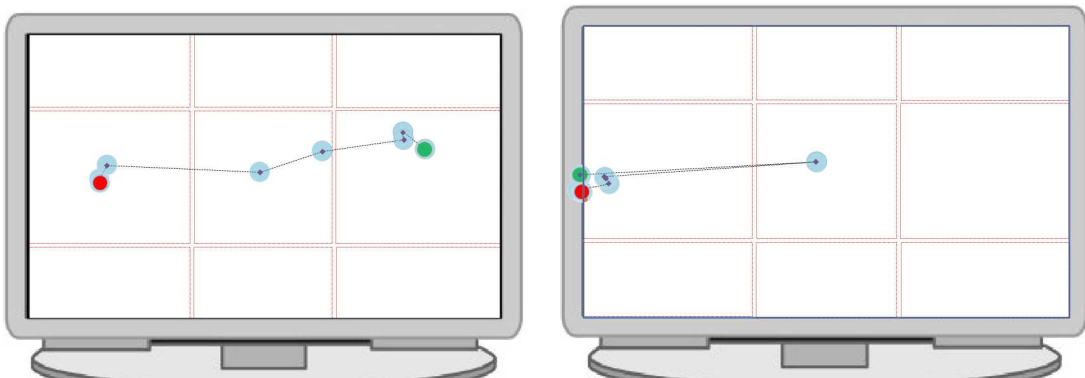


Figure 7.21: GP-A3: Fixation eye gaze scan path for zone 7 attempt 1 (left) and zone 8 attempt 2 (right). The green circle indicates the first fixation and the red circle the last fixation made.

the centre of the screen; then her gaze gaze following my finger back across the screen.

GP-A4, scored low $\overline{R_{user}}$ values for zone 3 (3%), 6 (19%), 7 (4%), 8 (8%) and 9 (19%). The failure to complete zones 3, 7 and 8 can be explained by the individuals eyes being lost (respectively) for 94%, 92% and 85% of the time. However for zones 6 and 9 the individual made prolonged fixations on one of the two test sequences. During zone 6 attempt 1, the eyes were lost for the majority of the time but during the second sequence the individual kept his gaze in one place (3 fixations recorded of up to 1234 ms each) and did not attempt to fixate within the correct zone, see Figure 7.22 (left). The same occurred during the zone 9 test sequences in which, the eyes were lost for 57% of

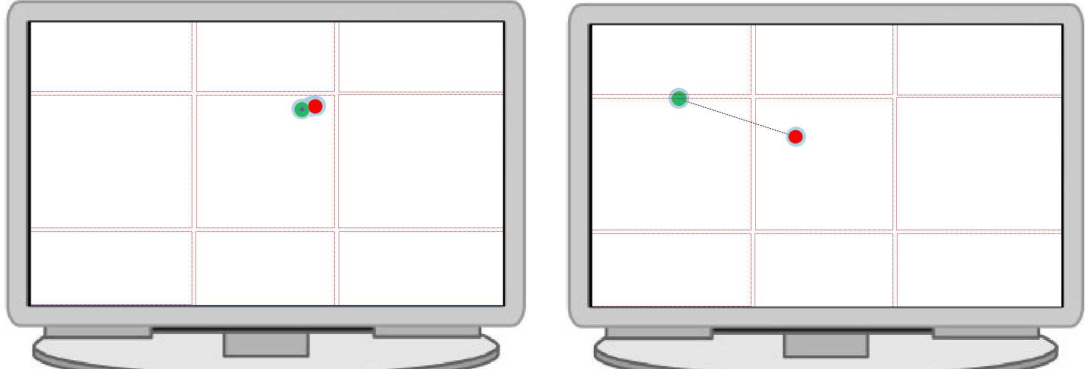


Figure 7.22: GP-A4: Fixation eye gaze scan path for zone 6 attempt 2 (left) and zone 9 attempt 1 (right). The green circle indicates the first fixation and the red circle the last fixation made.

the time during the first sequence but during the second sequence, the individuals gaze was kept in one place, see Figure 7.22 (right).

One possible explanation for both this user (GP-A4) and the previous (GP-A3) struggling with the bottom three zones could be due to them having difficulties in holding their head up right without support. Both participants struggle to support their head without a head rest and leaning forwards often results in their head falling forwards. In the zone 7, 8 and 9 test sequences the eye tracker has either lost the users eyes or they are fixating in the central part of the screen. If the individual is leaning their head forward when trying to hit these bottom zones then their head is likely to fall forwards, resulting in the eye tracker losing their eyes, and thus scoring poorly in the test sequence. Therefore, this suggests that these two participants may have chosen not to attempt some of these bottom zones as they knew their head would fall forwards and so wanted to keep their head up right.

7.6 Experiment 3: Validating the Adapted Interaction Technique

The observations and results from experiment 1 showed that the individuals in Group A had difficulties in using the current transparent overlay interaction technique. This resulted in poor control of their character in game. These difficulties were due to various

7.6 Experiment 3: Validating the Adapted Interaction Technique

influencing attributes categorised by the eye tracker hardware, environment and the individual themselves. Experiment 2 attempted to measure the influencing attributes for this particular interaction technique; eye tracker and individual at that moment in time, in the form of a reliability metric. $\overline{R_{user}}$ also suggested which elements of the interaction technique users were better at interacting with and thus allowed a modified version of the interaction technique to be created. The purpose of experiment 3 is to determine if the modified interaction technique allows for the influencing attributes and thus gives the individual better control over their character.

7.6.1 Configuration

World of Warcraft was used as the virtual environment. World of Warcraft and Snap Clutch were run from a Dell Latitude D830 laptop with 4GB of RAM and 1GB of Video RAM. An additional mode was developed for Snap Clutch that allowed the transparent overlay interaction technique to be adapted as required depending upon the diagnostic test results. Output from the laptop was sent to a 20 inch wide screen Samsung 206BW monitor set at a resolution of 1680x1050 px with a Tobii X120 eye tracker positioned below. The monitor and eye tracker were placed on a rise-and-fall table. The table height and eye tracker positioning angle was adjusted to suit each participant (as shown in Figure 7.3).

7.6.2 Procedure

This study was carried out with CP participants in Group A and followed the diagnostic test (experiment 2) after a short break of several minutes. The same eye tracker calibration used for the diagnostic test was applied to Snap Clutch, and the transparent overlay interaction technique was adapted as per the results. Following, Snap Clutch was started within the game and it was explained how the modified interaction technique for walking about in the world works. To aid the participants, the transparent overlays were visible at a low opacity level (15%). Each person was then free to walk around to familiarise themselves with the new configuration. The visible overlays were toggled on and off by myself and during this training phase they would be displayed and hidden until such time that each user was familiar. An example of how the visible overlays appear during this training phase can be seen in Figure 7.23.

7.6 Experiment 3: Validating the Adapted Interaction Technique



Figure 7.23: A modified version of the locomotion interaction technique. During this training phase, the transparent overlays are visible and can be switched on and off by the experimenter.

Once familiar, the participants were to complete an orienteering task within the game. Session notes were recorded as described in Section 7.3. Screen output was captured to file for determining task completion times and routes taken.

7.6.3 Orienteering Task

To evaluate the modified interaction technique for each user and to see if it has improved their character control in game, an orienteering challenge was devised. The idea being that each participant follows a navigational instruction to some landmark. Upon arriving at the landmark, a new navigational instruction is given. The World of Warcraft interface is created such that it can be easily modified using third-party software add-ons. TourGuide¹ is an add-on that allows you to create navigational way points on a map within the world. A map that had a combination of open spaces, buildings, trees and water was chosen for the task and three way points were created using the TourGuide add-on, see Figure 7.24 for a plan view of the map and way points. A second add-on was used to direct the participants in a GPS navigation manner: TomTom². TomTom uses the navigational way points from TourGuide and directs the users accordingly using a large green arrow; as they get within a certain range of the waypoint

¹TourGuide Add-on for World of Warcraft: <http://wow.curse.com/downloads/wow-addons/details/tourguide.aspx>

²TomTom Add-on for World of Warcraft: <http://wow.curse.com/downloads/wow-addons/details/tomtom.aspx>



Figure 7.24: Overhead view of the map with orienteering task way points overlaid. ‘1’ is the start point; ‘2’ is the first waypoint; ‘3’ is the second waypoint; ‘4’ is the final waypoint.

(10 in-game yards) then the waypoint updates to the next. Below TomTom’s navigational green arrow is a short description of where they need to head towards. The three way points and descriptions given were as follows:

1. 1st Waypoint: ‘Find the pond and swim to the centre’, Figure 7.25.
2. 2nd Waypoint: ‘Find the fort’, Figure 7.26.
3. 3rd Waypoint: ‘Find the centre of the village’, Figure 7.27 and Figure 7.28.

The way points were read to the participants as they appeared and they were free to take any route that they wished in order to reach them. None of the participants had prior knowledge to the way points or their locations. The time to reach each waypoint and the routes taken were recorded.

7.6.4 Analysis and Results

The aim of this experiment was to verify that the modified interaction technique allowed participants in Group A to have improved control over their character. This being based on a measure of reliability and so prior to examining the results our expectations are:

- Improvement in character control based on subjective evaluation when compared with experiment 1, see Table 7.2.

7.6 Experiment 3: Validating the Adapted Interaction Technique



Figure 7.25: Experiment 3: Start position of the orienteering task with the first waypoint given: ‘Find the pond and swim to the centre’. The green TomTom arrow moves dynamically as the character moves around in the world. When within 10 yards of the waypoint, it is updated with the next instruction.



Figure 7.26: Experiment 3: Waypoint 1 is reached and the location of the next waypoint is given: ‘Find the fort’.

7.6 Experiment 3: Validating the Adapted Interaction Technique



Figure 7.27: Experiment 3: Waypoint 2 is reached and the location of the final waypoint is given: 'Find the centre of the village'.



Figure 7.28: Experiment 3: Waypoint 3 is reached and the task is complete: 'Congratulations!!!'.

7.6 Experiment 3: Validating the Adapted Interaction Technique



Figure 7.29: GP-A2 performing the orienteering task in experiment 3.



Figure 7.30: GP-A3 performing the orienteering task in experiment 3.



Figure 7.31: GP-A4 performing the orienteering task in experiment 3.

- A correlation between the time taken to complete the orienteering task and $\overline{R_{user}}$

The ‘time taken to complete a task’ is a common measure of performance in HCI experimental settings. When a baseline of performance is established it offers a useful comparison. However, this may not be the most suitable measure for participants in Group A, as locomotion in virtual environments may still be a novel experience for them. Their priority may not be to complete a task as quickly as possible, despite being given instructions to do so. Therefore, it is important to also evaluate subjectively.

Waypoint Completion

GP-A2 used the unmodified version of the interaction technique and was able to navigate to all three way points, see Figure 7.32. Previously, she was able to navigate and control her character with little problem during experiment 1. GP-A1, GP-A3 and GP-A4 were all able to control their character in the world and follow navigation instructions to at least the first waypoint. This is a great improvement when comparing to the observations made during experiment 1 in which, all three individuals were only able to, at best, start and stop their character moving. However, despite having better control over their character they were unable to navigate to all way points. GP-A1 navigated to two way points with GP-A3 and GP-A4 only navigating to one. The difficulties observed during experiment 1 were still present with the modified technique,



Figure 7.32: Experiment 3: Route taken by GP-A2

that is for instance, the eye tracker losing their eyes; but there were now added difficulties that were previously experienced with able-bodied participants found in pilot studies 2 and 3 (see Chapter 5). On many occasions participants would turn too sharply and then try to re-correct their turning resulting in a turn-overshoot error. Also, individuals would get their character stuck against a fence or in a corner; this problem was particularly challenging to overcome with the modified technique as there was now no method for walking backwards. This resulted in these three participants becoming frustrated and thus making eye tracking even more difficult for themselves. However, it can also be considered positive that these difficulties are being experienced, as they demonstrate that the modified interaction technique has allowed them a comparable level of control to the able-bodied participants in the previous experiments. GP-A1 navigated his character into the fort after completing waypoint 2 but due to this being a small building he was unable to navigate out, see figure 7.33. Therefore, we brought the session to an end after two minutes of effort.

GP-A3 also got her character stuck in the small fort building whilst attempting to reach waypoint 1, see Figure 7.34. However, she was able to navigate back out of the fort and reach the first waypoint. During navigation to the second waypoint, the participant appeared to ignore the instructions and explore the areas of the world that she wanted to. Although, this has an affect on the remainder of the data collected for this part of the task for this participant, it is seen as a positive outcome as she was

7.6 Experiment 3: Validating the Adapted Interaction Technique



Figure 7.33: Experiment 3: Route taken by GP-A1



Figure 7.34: Experiment 3: Route taken by GP-A3

7.6 Experiment 3: Validating the Adapted Interaction Technique



Figure 7.35: Experiment 3: Route taken by GP-A4

<i>Participant</i>	Diagnostic Test		Time to Reach Way points (s)			
	$\overline{success}$	$\overline{R_{user}}$	<i>1st</i>	<i>2nd</i>	<i>3rd</i>	<i>Total</i>
GP-A1	94%	58%	86s	65s	DNF	DNF
GP-A2	94%	69%	43s	52s	33s	128s
GP-A3	55%	35%	75s	DNF	DNF	DNF
GP-A4	39%	34%	99s	DNF	DNF	DNF

Table 7.8: Experiment 3: Summary of task completion times and associated results from the diagnostic test, also see figure 7.36 (DNF = Did Not Finish)

able to independently go where she wanted. GP-A4 had many of the ‘turn-over shoot’ and ‘getting the character stuck’ difficulties when navigating to the first waypoint, see Figure 7.35. Upon reaching the first waypoint we brought the task to an end as he appeared frustrated and tired.

In terms of task completion time, GP-A2 was the only participant with a complete time: 128 seconds. Therefore, the time to reach only the first waypoint is the only time comparison between the four participants; of which GP-A2 was the quicker, Figure 7.8. When comparing the completion time to $\overline{R_{user}}$ and $\overline{success}$ there is no obvious correlation, Figure 7.36. GP-A1 and GP-A2 reached more way points and both achieved the highest $\overline{R_{user}}$ and $\overline{success}$ in comparison to GP-A3 and GP-A4 who achieved low values. However, GP-A1 time to reach waypoint 1 was similar to that of GP-A3 and GP-A4.

7.6 Experiment 3: Validating the Adapted Interaction Technique

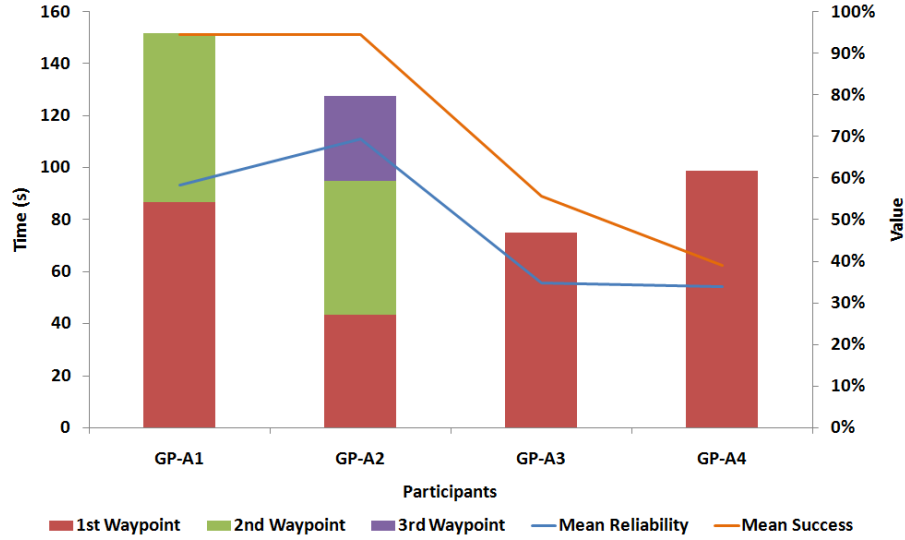


Figure 7.36: Experiment 3: Chart displaying the time to complete each waypoint and results from the diagnostic test ($\overline{success}$ and $\overline{R_{user}}$), also see Table 7.8

Participant	Experiment 1				Experiment 3			
	Task	Task	Task	Control	Task	Task	Task	Control
	1	2	3	Level	1	2	3	Level
GP-A1	Y	N	N	1	Y	Y	Y	5
GP-A2	Y	Y	Y	6	Y	Y	Y	6
GP-A3	Y	N	N	2	Y	Y	Y	5
GP-A4	Y	N	N	2	Y	Y	Y	4

Table 7.9: Experiment 3: Observation evaluation comparison between experiment 1 (highest observations) and 3. See Table 7.3 for definitions.

Observation Comparison

Upon comparing observations between experiment 1 and experiment 3 it can be seen that all individuals made an improvements in their ability to control their character movement, Table 7.9. Using the levels of character control found in Figure 7.9, GP-A1 has improved from level 1 to level 5; GP-A3 has improved from level 2 to level 5; GP-A4 has improved from level 2 to level 4; and GP-A2 has remained the same at level 6. Throughout the 3 parts to this experiment, there has been a large variance in the health of GP-A1. During experiment 1 his physical condition was poor and

subsequently performed poorly during both sessions for that first observation. His performance during experiment 2 and 3 was significantly better, as was his physical condition, although he did require some modification to the interaction technique.

7.7 Discussion and Conclusions

This Chapter has presented a three part evolving experiment with a focus on dynamic accessibility. The pilot studies performed with able-bodied participants in Chapter 5 and 6 were designed with accessibility in mind but not that it should be dynamic. The first understanding was that by developing eye-gaze only interaction techniques we would enable those with severe physical disabilities to be able to take part in immersive virtual environments. This was making an incorrect assumption of interaction techniques that work for able-bodied users will also work for disabled users. Dynamic accessibility is important because each individual has a unique set of needs and requirements and these may vary throughout the day depending upon physical and emotional states.

The three experiments in this Chapter involved four cerebral palsy participants that would benefit from eye-gaze as their primary communication device. The data collected was evaluated on a case study basis due to the low number of participants.

The first experiment demonstrated that the previously successful transparent overlay (Subsection 5.2.1), was not as successful for this group of individuals. One participant was able to use the technique and control their character well, where as the other three struggled and only able to perform a limited range of movements. However, despite the limited control capabilities the participants were engaged and motivated when playing the game. One individual completely relaxed his body when his character started to swim in the water suggesting that the environment and the characters current state has an affect on the user's own mental and physical state.

The second part of the experiment was an attempt to measure the individuals ability in using an interaction technique and if possible adapt the technique to better suit them. This was achieved by developing a diagnostic test that would function in a way similar to that of the interaction technique. A metric of reliability (R_{user}) was developed that assesses an individuals reliability in completing the diagnostic test rather than their success in completing it. This was based on the number of fixations made, the length

of time that the eyes were visible by the eye tracker, and the time taken to complete the test. Depending upon the individuals reliability for the entire test and with the different test sequences, the interaction technique can be modified using a priority system that assigns locomotion controls to different interaction areas on the screen. All four participants were able to complete the diagnostic test and the results showed that there were areas on the screen that they were able to look at better than others. Closer examination of gaze scan paths showed that on occasion individuals were distracted or lost interest in the sequence; or perhaps knew that they could not look well at parts of the screen so chose not to look at them. The age of the four participants ranged from 8 to 17 and so by changing the colours, using pictures or animated images rather than plain test buttons may have resulted in different results.

Based upon the results of the diagnostic test the interaction technique was modified and evaluated in the final part of the experiment. Three participants used a modified technique where as one did not due to her \overline{R}_{user} results in the diagnostic test being high. These three made significant improvements in their ability to control their character in the game, see Table 7.9. Due to these improvements, new problems in character locomotion were being experienced that were previously only experienced with able-bodied participants. This also verifies the improvement in character control with these participants being at a level of control seemingly comparable to the able-bodied participants. This modification was applied manually but both this and the diagnostic test could be automated and made as part of Snap Clutch. The process could also use machine learning methods that operates by collecting sets of diagnostic test data at different times of the day and then predicting which interaction technique configuration a person is likely to be best suited to.

The Chapter has described a series of experiments that have resulted in slightly steering away from one of the original objectives. That is, the direct selection and mapping of interaction technique to game task and the generation of a complete gaze user interface for use in a game. The first experiment demonstrated that it is important to first meet the range of needs for the disabled users that the interaction techniques are intended for. This resulted in a need to adapt techniques based on user ability of which, has been partly achieved. The next Chapter repeats experiments 2 and 3 with muscular dystrophy participants in group 2 and able-bodied participants in group 3.

Chapter deliverables:

- *An method that allows an interaction techniques to be customised to suit an individual user based on spatial positioning.*
- *Eye-gaze only interaction techniques developed and successfully validated with able-bodied participants does not guarantee that they will be suitable for use by severely disabled participants.*
- *Development of a reliability metric that when used within a specially designed diagnostic test, allows for an interaction technique to be adapted.*
- *The application of diagnostic tests as a means of assessing a user's reliability in performing an interaction technique.*
- *A visualisation model of viewing reliability based on 3D surface maps.*
- *An experiment that has validated, that the adapted interaction technique is better suited to the individual than the original.*

Chapter 8

Evaluation and Comparison Between all Group Participants

The previous Chapter presented a set of experiments that resulted in the transparent overlay interaction technique being adapted to suit cerebral palsy (CP) participants in user Group A based upon dynamic accessibility. This involved performing a diagnostic test that was measured by a metric of reliability (R_{user}); the technique adapting based upon the results; and finally verification that the adapted technique works by performing an orienteering task in World of Warcraft. This Chapter repeats the diagnostic test and the orienteering task with muscular dystrophy (MD) participants in user Group B and able-bodied (AB) participants in Group C.

8.1 Experiment 4: Determining User Reliability (R_{user})

Experiment 2 in Section 7.5 measured R_{user} for each participant and subsequently determined if the interaction technique should be adapted or not. The purpose of this experiment is to repeat this but with MD and AB participants.

8.1.1 Participants and Configuration

The same diagnostic test found in Section 7.5 was used for this experiment. The software was run from a Dell Latitude D830 laptop with 4GB of RAM and 1GB of Video RAM. Output from the laptop was sent to a 20 inch wide screen Samsung

8.1 Experiment 4: Determining User Reliability (R_{user})

monitor set at a resolution of 1680x1050px with a Tobii X120 eye tracker positioned below.

There were eight participants from Group B and they were all year 12 or 13 students at Ash Field school, aged from 15 to 18 years with (mean = 16 years). The session took place at the school for six of the participants and at the eye-gaze lab at De Montfort University for the remaining two. Identical equipment was used for all sessions. A height adjustable table was used and eye tracker positioning angle was adjusted to suit each participant, Figure 7.3 shows a typical arrangement for both sessions. All had previously used an eye tracker at least once but were all inexperienced users.

All twenty participants from Group C either worked or studied at De Montfort University; 16 were male and four were female. They were aged 20 to 58 years with (mean = 30 years). Seven participants had used an eye tracker at least once previously with six being inexperienced and one being an expert user.

8.1.2 Procedure

An explanation of how the tests worked was given and each individual was allowed to run through part of the test to facilitate understanding. Following, the test began as described in Section 7.4.2.

8.1.3 Results and Analysis

All participants were able to complete the test. In comparison with the results for CP participants in Group A; the median R_{user} for the MD participants was 81%, for AB participants was 86%, with the CP participants being 46%, see Table 8.1 and Figure 8.1.

	n	<i>Median</i>	<i>Mean</i>	σ
Group A	4	46%	49%	15%
Group B	8	81%	78%	12%
Group C	20	86%	87%	4%

Table 8.1: Reliability (R_{user}) results for all three participant groups. Group A contains CP participants; Group B contains MD participants; Group C contains AB participants.

8.1 Experiment 4: Determining User Reliability (R_{user})

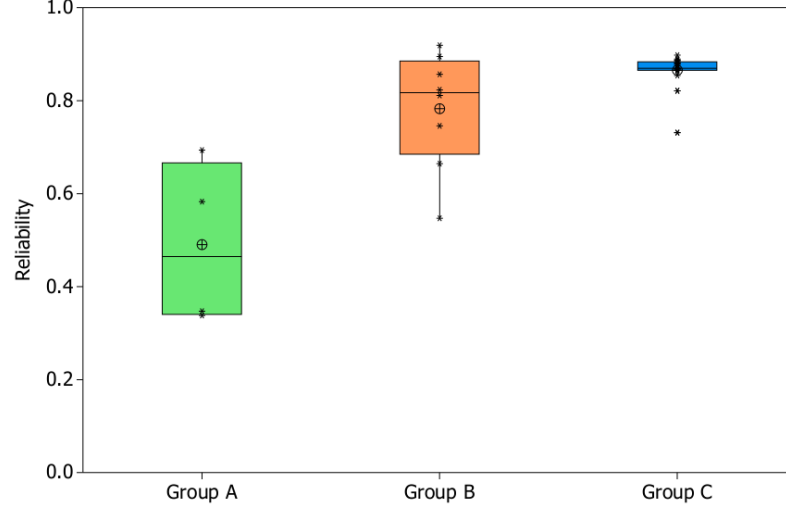


Figure 8.1: Experiment 4: Reliability by user group. Group A contains CP participants; Group B contains MD participants; Group C contains AB participants.

By examining the interquartile range of the AB group, it can be seen that there is a tight distribution with a high level of R_{user} and only few outliers. Whereas, there is a wider distribution with the CP and MD groups; although, the MD group level of R_{user} was close to that of the AB group.

Taking the four CP participants in Group A and plotting their R_{user} values of 34%, 35%, 58% and 69% onto a histogram of Group B and Group C values, it can be seen that all four points sit outside the range of Group C and two points outside the range of Group B, see Figure 8.2. This suggests that there is a significant difference between participants in Group A and Group C and participants in Group A and Group B. Re-sampling¹ methods were used to verify the significance (see Appendix B for re-sampling notes and all data collected for experiment 4):

- CP participants and MD participants = 75 out of 10,000 ($p = 0.0075$) of the re-sampled statistics are more or less extreme than our observed result so we can reject the null hypothesis²
- CP participants and AB participants = 0 out of 10,000 of the re-sampled statistics are more or less extreme than our observed result so we can reject the null

¹Data was re-sampled into groups 10,000 times without replacement.

²Null hypothesis being all samples drawn from one population.

8.1 Experiment 4: Determining User Reliability (R_{user})

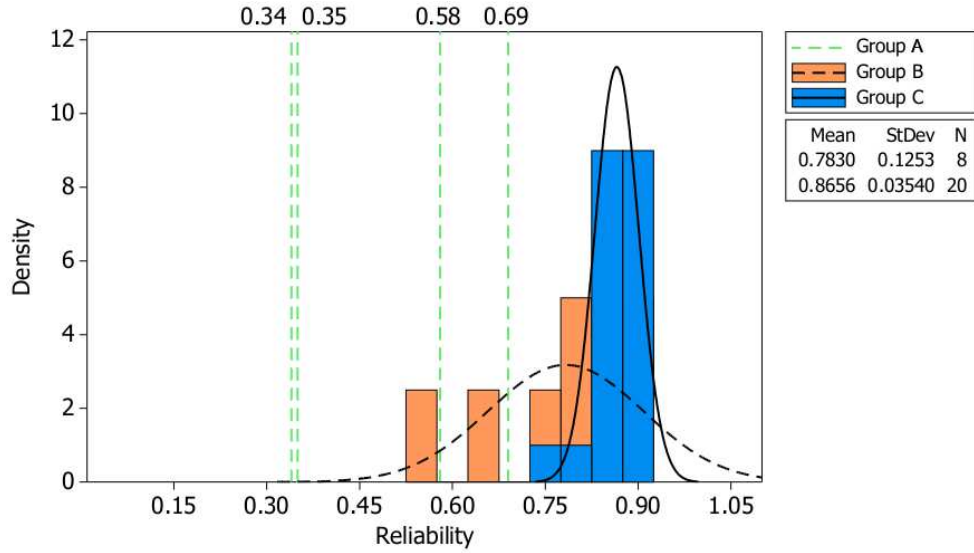


Figure 8.2: Experiment 4: Reliability distribution of Groups B (MD) and Group C (AB). The vertical green lines indicate the reliability values for CP participants in Group A. The chart shows that all four participants sit outside the distribution of Group C and two outside the distribution of Group B.

hypothesis

- MD participants and AB participants = 61 out of 10,000 ($p = 0.0061$) of the re-sampled statistics are more or less extreme than our observed result so we can reject the null hypothesis

By looking closer at the fixation component of the reliability metric it can be seen that the more fixations an individual makes then the longer it takes them to complete the test, see Figure 8.3. The data points for the AB participants in Group C have a much tighter distribution than the CP (Group A) and MD (Group B) groups. This can also be seen when examining the length of time the eyes are not being tracked by the eye tracker, see Figure 8.4.

Figures 8.5, 8.6 and 8.7 show the distribution of fixations, completion time and length of time eyes are tracked for all three groups. In all of these attributes, the CP participants have a much wider distribution than the AB and MD participants. Examining only fixations (Figure 8.5), the AB and MD participants have similar distributions. This suggests that AB participants may be used to represent MD participants when there is a requirement to measure based upon fixation count. The completion

8.1 Experiment 4: Determining User Reliability (R_{user})

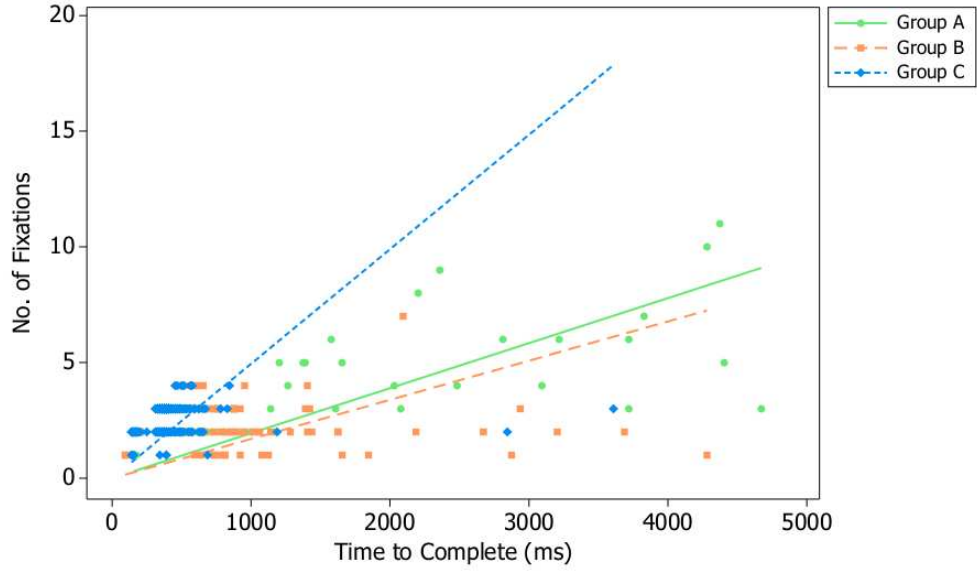


Figure 8.3: Reliability metric/diagnostic test: Number of fixations made versus time to complete a test sequence. Group A contains CP participants; Group B contains MD participants; Group C contains AB participants.

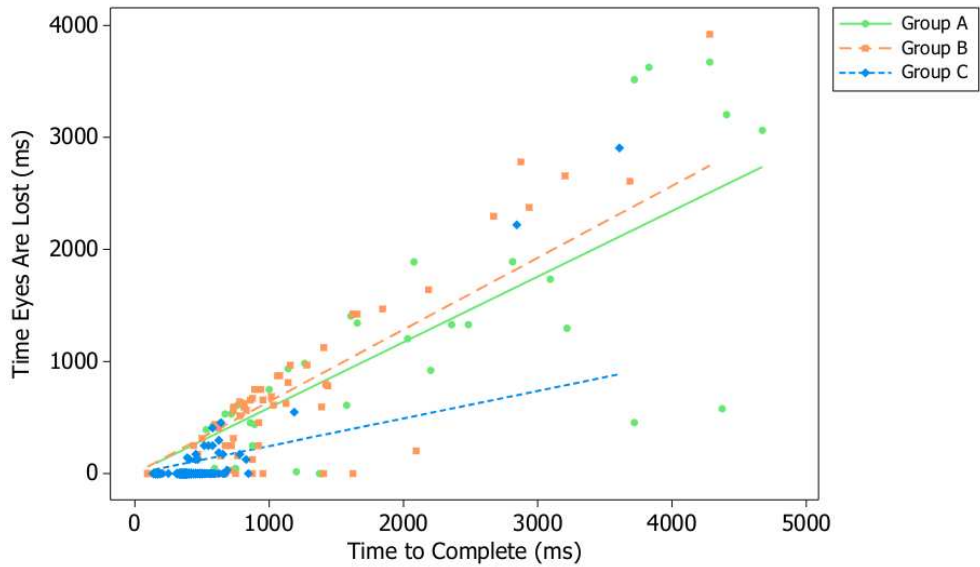


Figure 8.4: Reliability metric/diagnostic test: Length of time the eyes were not visible by the eye tracker versus time to complete a test sequence. Group A contains CP participants; Group B contains MD participants; Group C contains AB participants.

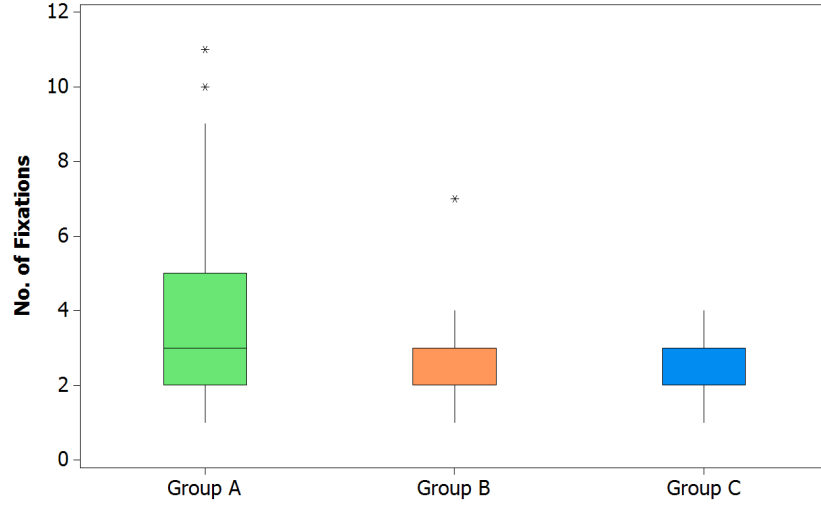


Figure 8.5: Distribution of fixations made for CP participants (Group A), MD participants (Group B) and AB participants (Group C)

time (Figure 8.6) is much tighter for AB participants than MD participants. The difference in distribution between groups appears much greater when examining the length of time the eyes are not tracked (Figure 8.7) with AB participants being tracked almost all of the time. There is a general assumption that those with severe disabilities often retain good control over their eye movement however, the observed data shows that this is not necessarily the case for those suffering from CP. There is an argument that with practice, the results may improve however, this would require a further longitudinal study.

In summary:

- *The observed reliability (R_{user}) of AB participants is significantly higher and more tightly distributed than MD participants and CP participants.*
- *The observed reliability (R_{user}) of MD participants is significantly higher and more tightly distributed than CP participants.*

8.2 Experiment 5: Locomotion Task

Similar to experiment 3, the purpose of this experiment is to verify that the proposed interaction technique (albeit modified or not) allows participants to perform a loco-

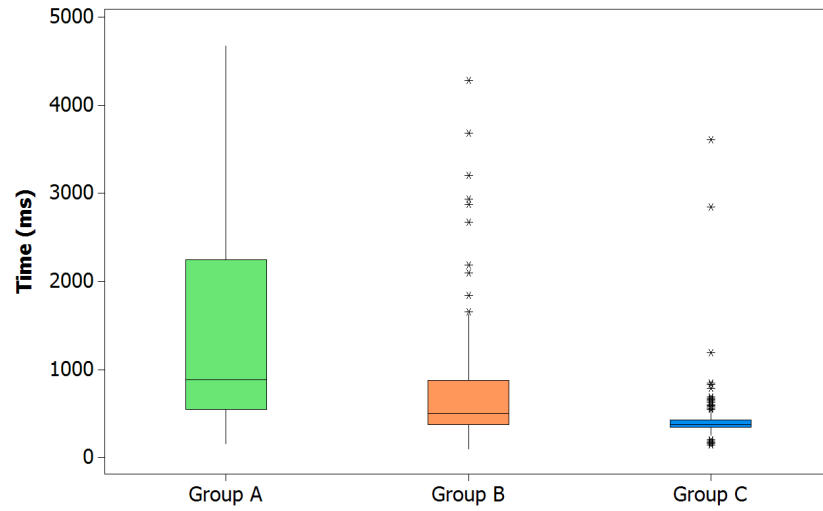


Figure 8.6: Distribution of completion time for CP participants (Group A), MD participants (Group B) and AB participants (Group C)

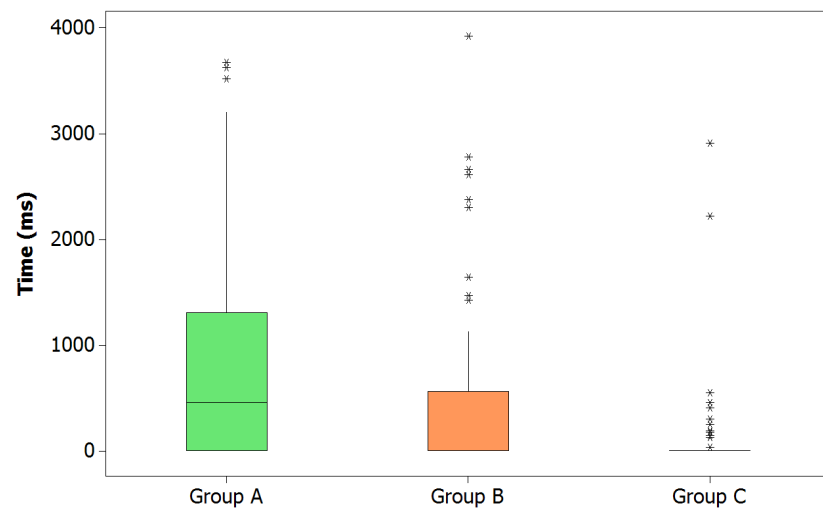


Figure 8.7: Distribution of time eyes are not being tracked for CP participants (Group A), MD participants (Group B) and AB participants (Group C)

tion based orienteering task in a virtual environment.

The results of the diagnostic test in experiment 4 suggest that all AB participants in Group C and six MD participants from Group B do not require any modification to the transparent overlay interaction technique. Only two MD participants in Group B achieved low R_{user} requiring some modification; GP-B2 and GP-B3. Using the priority function system, their R_{user} surface map and adapted interaction technique appears as Figure 8.8 and Figure 8.9.

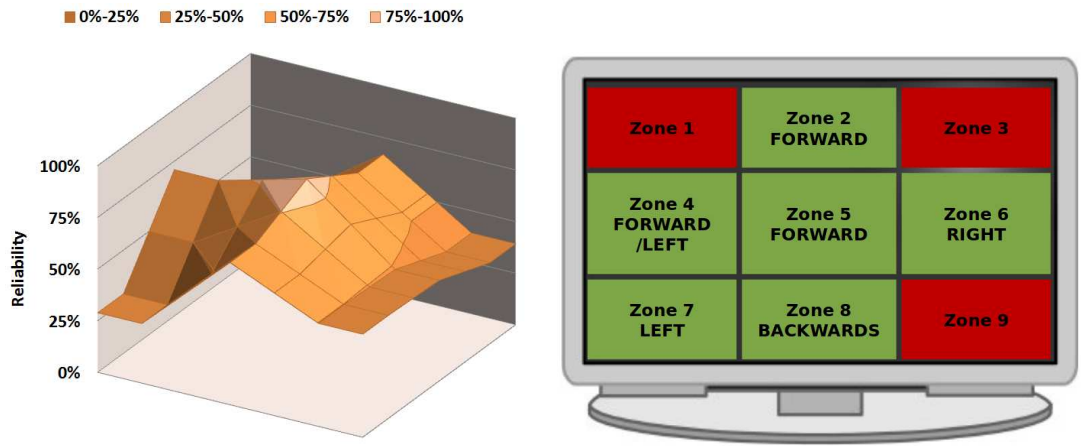


Figure 8.8: Experiment 5: MD participant GP-B2' reliability surface map (left) and their adapted interaction technique (right).

8.2.1 Participants and Configuration

World of Warcraft was used as the virtual environment. World of Warcraft and Snap Clutch were run from a Dell Latitude D830 laptop with 4GB of RAM and 1GB of Video RAM. An additional mode was developed for Snap Clutch that allowed the transparent overlay interaction technique to be adapted as required depending upon the diagnostic test results. Output from the laptop was sent to a 20 inch wide screen Samsung 206BW monitor set at a resolution of 1680x1050 px with a Tobii X120 eye tracker positioned below.

This experiment followed experiment 4 after a short break of several minutes. All of the participants had taken part in experiment 4. One of the AB participants from Group C was not feeling well leaving nineteen participants from this group. All eight MD participants from Group B took part.

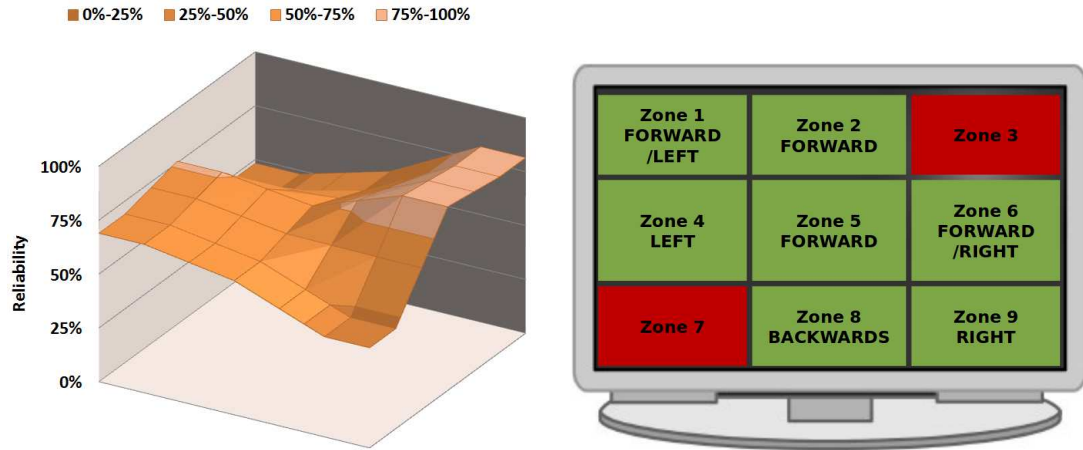


Figure 8.9: Experiment 5: MD Participant GP-B3's reliability surface map (left) and their adapted interaction technique (right).

8.2.2 Procedure

The same eye tracker calibration that was used for the diagnostic test was applied to Snap Clutch, and the transparent overlay interaction technique modified as per the results. Following, Snap Clutch was started within the game and an explanation of how the interaction technique worked was given. Each participant was then free to walk around in the world to familiarise themselves with the technique. If the technique had been modified, then visible overlays were toggled on and off until the individual was familiar. Following this, it was explained that the participants would be performing an orienteering task in which, they would be required to navigate to three different points. The orienteering task was the same as used in experiment 3, see Subsection 7.6.3. Screen output was captured to file for determining task completion times and routes taken.

8.2.3 Results and Analysis

All participants were able to reach the first waypoint with 89% of the AB participants and 88% of the MD participants reaching all three way points, see Table C.1 in Appendix C.

Two AB participants (GP-C18 and GP-C20) had difficulty in using the interaction technique and commented how it was difficult to keep their head still when trying to

turn the character. At the start of the training phase all participants were told to keep their head as still as possible and to only move their eyes. Therefore, a possible explanation is that they were able to keep their head still at the start of the session and able to reach the first waypoint in a reasonable time. Then, as they settled into the experiment they began to move their head more, thus losing some control of their character.

The MD participant in Group B (GP-B2) that was only able to reach the first waypoint struggled in turning their character around and became stuck in some trees and bushes. As this individual was using a modified version of the interaction technique they found it more difficult to manoeuvre the character as was required. GP-B3 also used a modified version of the interaction technique but was able to reach all three waypoints, albeit within a much longer time than other participants in Group B.

First Waypoint Comparison for all Participants

As all AB, MD and CP participants from the three groups were able to reach at least the first waypoint, it is more appropriate to use only the first waypoint time data for comparison of all groups.

When comparing the results to Group A in experiment 3; the median time for Group B was 35 seconds, for Group C was 31 seconds, with Group A being 81 seconds, Table 8.2 and Figure 8.10.

	n	<i>Median</i>	<i>Mean</i>	σ
Group A	4	80.8 sec	76 sec	20.7 sec
Group B	8	35.2 sec	35.8 sec	7.5 sec
Group C	20	30.5 sec	31 sec	9.2 sec

Table 8.2: Experiment 5: Time for all three participant groups. Group A contains CP participants; Group B contains MD participants; Group C contains AB participants.

The order of group task time performance is the same as with the reliability results: CP Group A < MD Group B < AB Group C. However, the difference between the MD group and the AB group appears much smaller than with the reliability results, with the distribution for the MD group similar to that of AB but with the AB group being slightly tighter and with a quicker median time. There still remains a large difference

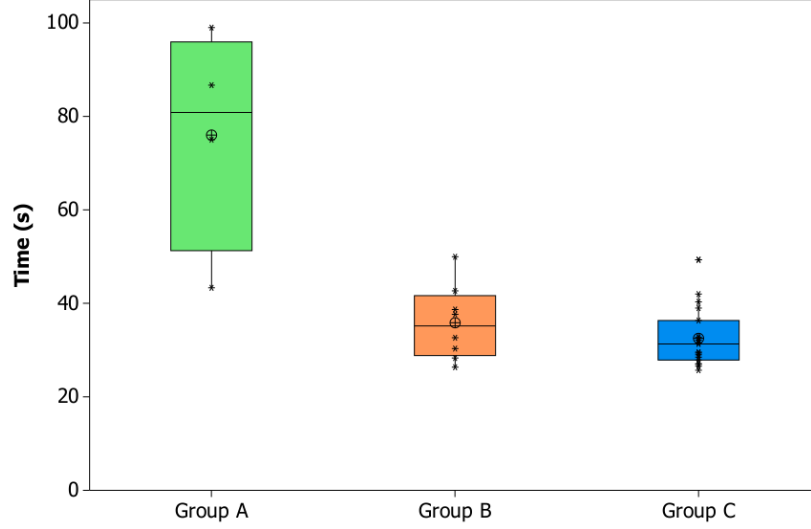


Figure 8.10: Experiment 5: Time to reach the first waypoint by user group

when comparing both AB and MD groups with the CP group, with the CP group having a much wider distribution and a longer median task time.

Taking the four CP participants in Group A and plotting their task times of 87, 43, 75 and 99 seconds onto a histogram of MD and AB participant values, it can be seen that only one CP participant achieves a similar task completion time, see Figure 8.11. This individual (GP-A2) was the only participant in Group A that was able to use the interaction technique unmodified and also achieved a high reliability rating. This Figure suggests that there is a significant difference between CP participants in Group A and AB participants in Group C; and CP participants in Group A and MD participants in Group B. Re-sampling¹ methods were used to verify the significance (see Appendix C for re-sampling notes and all data collected for experiment 5):

- CP participants and MD participants = 94 out of 10,000 ($p = 0.0094$) of the re-sampled statistics are more or less extreme than our result so we can reject the null hypothesis
- CP participants and AB participants = 8 out of 10,000 ($p = 0.0008$) of the re-sampled statistics are more or less extreme than our result so we can reject the null hypothesis

¹Data was re-sampled into groups 10,000 times without replacement.

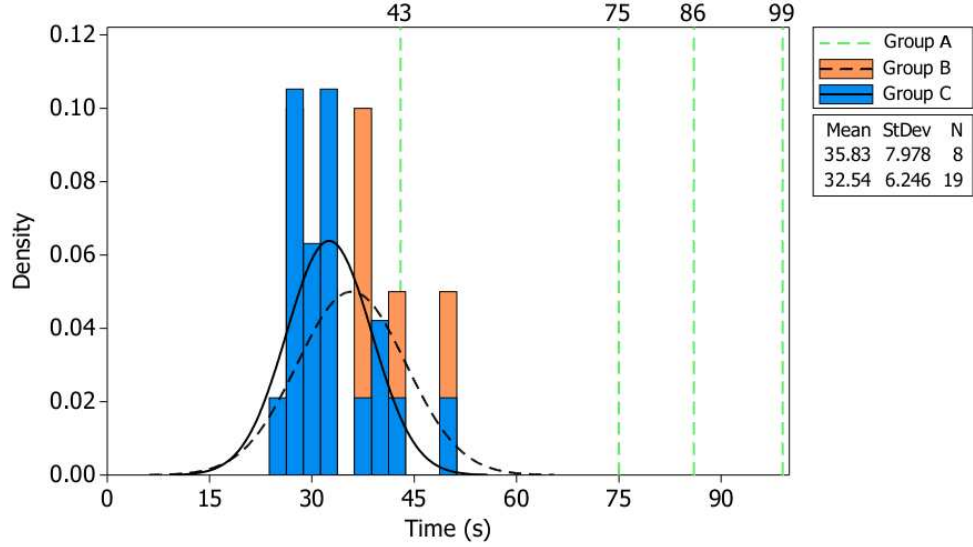


Figure 8.11: Experiment 5: Time to reach first waypoint, distribution of Group B and Group C. The vertical green lines indicate the time for participants in Group A. Group A contains CP participants; Group B contains MD participants; Group C contains AB participants.

- MD participants and AB participants = 1760 out of 10,000 ($p = 0.176$) of the re-sampled statistics are more or less extreme than our result so we cannot reject the null hypothesis

Complete Waypoint Comparison for Able Bodied (AB) and Muscular Dystrophy (MD) Participants

There was seven MD participants from Group B and seventeen AB participants from Group C that were able to reach all three waypoints. A summary of the results can be seen in Table 8.3 and in Figure 8.12.

	<i>n</i>	<i>Median</i>	<i>Mean</i>	σ
Group B	7	83 sec	91.7 sec	21.8 sec
Group C	17	81 sec	84.5 sec	15.8 sec

Table 8.3: Experiment 5: Total time for participant Groups B and C. Group B contains MD participants; Group C contains AB participants.

The mean and median values indicate that although the AB participants performed

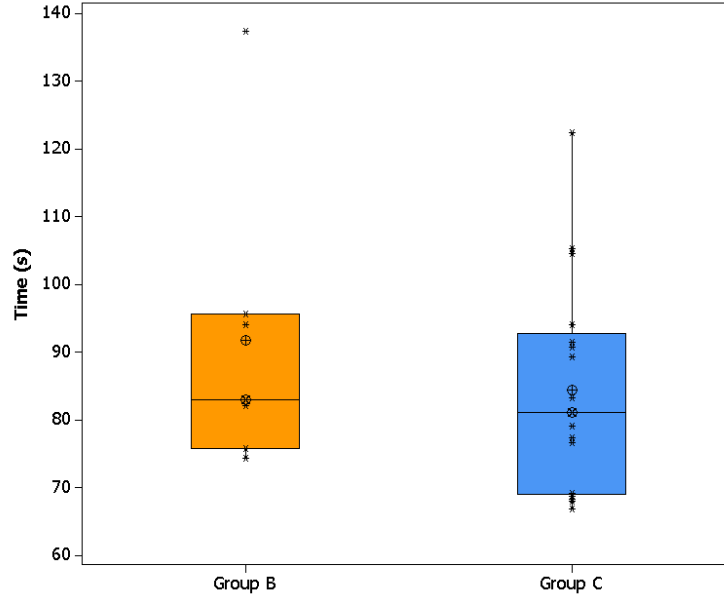


Figure 8.12: Experiment 5: Total time to reach all waypoints for MD participants in Group B and AB participants in Group C

the task in a quicker time, they are very close to the values achieved by the MD participants. Re-sampling¹ methods were used to verify the significance (see Appendix C for re-sampling notes and all data collected for experiment 5):

- MD participants and AB participants = 7,097 out of 10,000 ($p = 0.71$) of the re-sampled statistics are more or less extreme than our result so we cannot reject the null hypothesis.

This supports the previous comparison between these two groups in that there was no observed significant difference in locomotion task performance.

In summary:

- *The difference observed in locomotion task time performance is significant for CP participants in Group A when compared to MD and AB participants in Group B and Group C.*
- *The difference observed is not significant for MD and AB participants in Group B and Group C.*

¹Data was re-sampled into groups 10,000 times without replacement.

8.3 Discussion and Conclusions

This Chapter has presented two experiments with MD and AB participants from Group B and Group C. The results were compared to those obtained from CP participants in Group A in experiment 2 and experiment 3 (found in the previous Chapter).

The first experiment was a diagnostic test identical to the one featured in experiment 2. The aim of this was to determine the user reliability (R_{user}), with the expectation being that the AB group would achieve a higher R_{user} than the MD group; and the MD group higher than the CP group. This was found to be true when comparing the median R_{user} and also on an individual basis when plotting the results from the four CP participants onto a histogram of AB and MD group distributions (Figure 8.2). This difference was also found to be significant when applying re-sampling statistical methods.

The second experiment required the participants to navigate and control their character in a virtual environment. This was an identical task exercise to the one used for experiment 3. The results from the first experiment indicated that two MD participants required a modification to the interaction technique. All group participants were able to navigate to at least one waypoint, with two AB participants from Group C who had difficulties due to head movement and losing control of their character; and one MD participant from Group B who became stuck in some trees and bushes whilst using a modified version of the technique. In comparing the task times to reach the first waypoint, the AB group performed better than the MD group who performed better than the CP group.

In addition to the individuals disability, another discussion for the lower R_{user} and increased locomotion task time with CP participants and the better with MD participants, is that the CP participants have been limited to access via switches and only recently introduced to eye-gaze. This results in a lack of stimulation through not being able to access the same resources, self-expression and level of interaction. All of the MD participants are able to communicate verbally and use a wider range of input devices allowing for greater stimulation. The AB participants go a step further with all participants experiencing no restriction to resources, interaction or self-expression. This is an affect of the emotional and development aspect of the disability rather than the physical disability itself.

This comparison however, does not apply when it comes to examining the locomotion task times of AB and MD participants in Group B and Group C. Here, the task times for the MD group were not significantly different from the AB group. This observation suggests that results from AB participants may be a good representation of MD participants in terms of this type of locomotion experiment.

One expectation is that the locomotion task times will be reflected by the individual's R_{user} ; *as the task time increases the reliability decreases*. This is true when comparing median values at group level, see Table 8.4. A Spearmans rank correlation

	Diagnostic Test	Locomotion Task
	<i>Median Reliability (R_{user})</i>	<i>Median Task Time</i>
Group A	46%	81 seconds
Group B	81%	35 seconds
Group C	86%	31 seconds

Table 8.4: The greater the reliability then the better task time performance when sorting by group median. Group A contains CP participants; Group B contains MD participants; Group C contains AB participants.

coefficient suggests a moderate negative correlation¹ of $\rho = -0.63$, see Figure 8.13.

However, there are several possible factors that may affect the correlation, such as:

- The diagnostic test not being as engaging as the locomotion task
- Although the test behaves similar to the interaction technique it does simulate the use of the technique whilst in an immersive environment, i.e. there are many distractions in the virtual world oppose to a distraction-free controlled test
- The factors that define the reliability metric (number of fixations, time taken, time eyes are present) are given equal rankings rather than weighted rankings
- Participants became more relaxed as the session progressed thus performing better after the diagnostic test
- Participants are anxious when taking a test but relaxed when playing a game

¹A negative correlation implies the direction of the correlation.

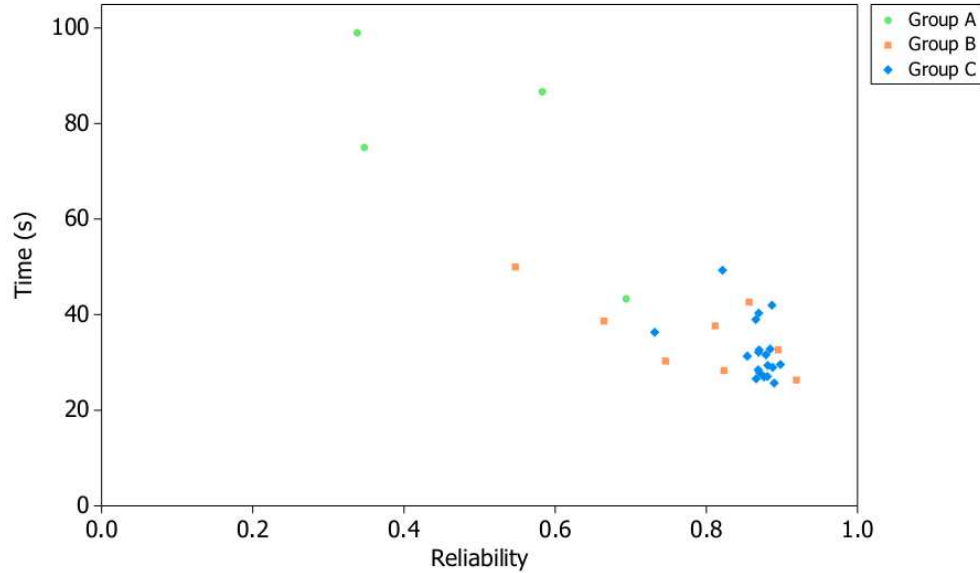


Figure 8.13: Time to reach the first waypoint versus reliability. No linear regression lines are shown as the data has been analysed using rank correlation. The greater the reliability then the better task time performance when sorting by group median.

Despite, the negative correlation between reliability and task time being only moderate, it's use does demonstrate how dynamic accessibility can be applied by customisation of an interaction technique. Thus allowing locomotion control when previously there was little or none. However, the affecting factors above need to be considered further before assuming a definitive correlation between the two variables.

Chapter deliverables:

- *Data from two experiments suggesting that there is a significant difference in locomotion task performance between cerebral palsy participants and able-bodied and muscular dystrophy participants.*
- *Further, the difference in locomotion task performance is not significant between muscular dystrophy participants and able-bodied. This observation suggests that results from able-bodied participants may also reflect results from muscular dystrophy participants.*
- *Finally, the experiments validate the reliability metric (R_{user}) in that, the greater the reliability then the greater the locomotion task time.*

Chapter 9

Conclusions

The aim of this work was to develop a means of selecting the most appropriate gaze interaction techniques for a set of specific game tasks. The intention being that a complete gaze interface could be constructed that would allow a user access to the full range of game tasks. This Chapter summarises this work and suggests areas of future research.

9.1 Contributions

The first part of this thesis provides background information into eye tracking and reviews previous work on gaze interaction. Understanding how the eye is constructed and moves is essential to understanding the technology. The intended end-user of this work, are those who suffer from a severe physical disability. To review all disabilities goes beyond the scope of this thesis but it is important to give a broad description so as to understand the nature of certain conditions. In particular, neuromuscular conditions such as muscular dystrophy and neurological conditions such as cerebral palsy. Although, low-tech and high-tech augmentative and alternative communication (AAC) devices exist they are all limited in terms of the available bandwidth that they offer. Eye and gaze tracking is a high-bandwidth AAC device that may be used as an alternative. However, there are many implications that can affect its use for people with physical disabilities, such as eye anatomy, eye conditions such as nystagmus, eye tracking technology and finally body positioning and involuntary movement.

The second part of the thesis addresses the research question of game tasks being matched with candidate gaze interaction techniques. The main contributions are:

- *A collection of novel gaze based interaction techniques that have been evaluated through empirical studies.*
- *A methodology for performing task analysis based on expert gameplay using gaze data, low-level event signature modelling and think-aloud protocols.*
- *An extensible software architecture that encompasses all of the novel gaze interaction techniques presented in the thesis. This has been made available for free download.*

This part began with an initial assessment of gaze based mouse emulation in performing standard virtual environment tasks (*Chapter 4*). Taking Second Life as an example of a Multi-User Virtual Environment (MUVE) or Massively Multiplayer Online Game (MMOG) a pilot study was conducted. The results helped to define several problems or errors that became common throughout the remainder of the experiments found in the thesis namely: path deviation, distraction, selection, accuracy, feedback. Although, task times were poor and error rates high it was possible to complete all five tasks using gaze based mouse emulation.

Following, a bottom-up approach of development and evaluation of gaze interaction techniques (*Chapter 5*) was conducted in parallel to a top-down analysis of expert gameplay (*Chapter 6*). The bottom-up approach attempted to develop and describe different interaction techniques in terms of the five different tasks found in the pilot study. These were implemented into middle-ware based software that is independent of the eye tracker and the target application. The intention being that it can be used with any game that accepts input events direct from the Windows message queue. The possibility of modifying and adapting the techniques was also explored and a taxonomy of gaze technique modification was developed. Two evaluation studies demonstrated that a basic level of gameplay using only eye-gaze was possible by able-bodied participants.

The top-down analysis presented a methodology for performing a detailed task analysis of a game and matching of candidate interaction techniques with game tasks. This was based on expert task analysis in which, gaze position, low-level event data and

think-aloud data was recorded when participants were asked to perform a typical in-game quest. All of the data was annotated into the sub-tasks that were performed and this showed the visual areas of interest (AOIs) and the low-level input events required for each sub-task. Using existing modelling methods as a starting point, a visualisation model based on low-level event data was created that allowed us to verify that a locomotion task using gaze interaction operated in a way similar to that of mouse and keyboard. Further attempts on modelling were also explored but these became much too complex. However, the analysis did allow for candidate interaction techniques to be suggested for different game tasks.

The final part of the thesis was to evaluate the interaction techniques with the intended end user. The main contributions are:

- *An investigation into the notion of dynamic accessibility.*
- *The application of diagnostic tests as a means to assess a person's reliability in performing a specific interaction technique. This involved the definition of a reliability metric and a method of modelling reliability based on 3D surface maps.*
- *A method of adapting an interaction technique based upon the diagnostic test results and reliability model that results in an improvement in task performance.*
- *Reliability and task performance evaluations that allowed generalisations about differences between different disability groups to be made.*

With the help of a specialist disability school, a group of children aged from 8 to 17 were selected to participate in a series of trials. The children fell into two groups: those who would greatly benefit from eye-gaze at this moment in time and those that would likely benefit from eye-gaze in the future. The first group consisted of four cerebral palsy (CP) sufferers (Group A) and the second group, eight muscular dystrophy (MD) sufferers (Group B). The novelty with this approach is that attempts are made to quantify differences between groups of people with different disabilities. This positivist based approach contrasts greatly with the phenomenalist approach of Donegan et al. in which there is little or no generalisation made and each person is considered as being unique. A third group was created made of twenty able-bodied (AB) participants

(Group C). This group was to act as a benchmark to compare the two disability groups against.

The first experiment (*Chapter 7*) demonstrated that all of the interaction techniques presented in the previous chapters of the thesis, that were designed for use by disabled users, did not work as intended with participants with physical disabilities. The common belief amongst many who research in this field has always been that those with such disabilities maintain good control over their eye movements. However, the nature of the individuals condition meant that the performance of the eye tracking used in the experiments was not consistent enough in tracking their eyes. This was mainly due to involuntary head movements, eyes being obscured by head rests, head falling forward whilst looking at the bottom of the screen, the individuals health at the time and so on. The original intention of this work was to select candidate gaze interaction techniques for a set of specific game tasks. Before this can take place however, we first need a mechanism of adapting the existing techniques to suit an individual. This first experiment thus provides the basis for dynamic accessibility in that levels of accessibility are unique to each user and these need to be able to adapt accordingly. The premise being that a user's requirements may alter depending upon their specific disability, at a certain time of the day, and when they are feeling a specific way emotionally and physically.

Using the influencing factors on gaze interaction that were observed during the first experiment, a diagnostic test was devised that was to assess an individuals reliability in using a specific interaction technique. The metric used to assess reliability was based on three factors which were: the time taken to complete each step of the test, the proportion of time that the eyes were being tracked, the number of fixations made during each step. Taking the locomotion task and the transparent overlay interaction technique as an example the second experiment evaluated the metric for each user. This was represented using a 3D surface model that highlighted the areas on the screen where the interaction technique was stronger or weaker for each individual. This allowed the technique to be customised (based on spatial positioning) to better suit the individual. Although, the adaption applied was simple it produced improvements in performance, as found in the third experiment. This third experiment was a structured orienteering task in World of Warcraft in which, participants were asked to follow a GPS style navigational arrow to three different way points. Here, three out of four of

the participants improved on their ability to perform the task that the interaction technique was designed for. The fourth participant was able to use the original interaction techniques from the outset without the need for modification.

The fourth and fifth experiments (*Chapter 8*) repeated the diagnostic test and orienteering task with the other two groups. The hypothesis for the experiments was that the AB participants would perform better than the MD participants who would perform better than the CP participants. This was found to be true for both experiments however, what was not expected is that there was a significantly higher difference in reliability between the two disability groups. Further, although there was a significant difference between the CP participants and the other two groups for the orienteering task, there was no significant difference between the MD group and the AB group. This also suggests that trials with AB participants can be representative of MD participants. Of course, MD is a very different type of condition to CP and in time their reliability and task performance may decline, but this further supports the case for providing dynamic accessibility.

There was a correlation expectation between the reliability metric and task performance time and when comparing median group values the indication is that the greater the reliability (the diagnostic test) then the faster the task time (the locomotion task). However, a Spearmans rank correlation suggests that it be only moderate ($\rho = -0.63$). There are a number of possible reasons for this such as test anxiety, weighting of the reliability metric factors, diagnostic test not being engaging enough and so on. Despite this correlation not being as high as was hoped, the reliability metric and diagnostic test provided adequate information to allow the interaction technique to be modified enough so that an improvement in locomotion task performance could be made.

9.2 Critical Review and Future Work

The research question posed at the start of the thesis (see Section 1.2) was:

Can an eye-gaze only interface be automatically generated to suit the specific needs of a person with severe physical disabilities so that they may fully participate in an immersive computer game?

At the start of this work it was expected that if an eye-gaze only interaction technique works for an able-bodied user then it would also work for a disabled user. This was

an incorrect assumption and as such the question became more difficult to answer. As such, the direction of the research changed to explore why the interaction techniques do not work as expected and how they can be customised. Earlier in this Chapter it was mentioned that one of the novelties in the methodology used was that this work was based on a positivism approach. This could be considered as contradicting in that the work also suggests customisation and dynamic accessibility to suit an individual. However, expert knowledge would normally be required for that customisation to take place; someone to analyse a new set of game tasks and then a programmer to build a new gaze interaction technique that the individual could use. By making some generalisation we can begin to remove the need for some of the expert knowledge. Gajos was able to achieve automatic generation of dialog boxes within standard desktop applications based on elicitation and diagnostic tests. In contrast to games, the tasks in these applications are always based upon pointing, dragging, list selection and clicking. Further, there is no real-time element to task completion. However, this does demonstrate the feasibility of automatic generation of a user interface according to user needs. Although, elicitation based questioning with young, non-verbal participants is difficult. The work in this thesis has shown that novel gaze interaction techniques can be created, they can be matched to game tasks and they can be customised to suit an individual. This process is not yet automatic but with further research and development this can be implemented using methods such as heuristics, elicitation and computational intelligence. The belief is that an individual who has purchased a new game, could perform several short diagnostic tests and as a result, a set of customised interaction techniques be generated that would allow him to play. Depending up on the user, the metric could alter drastically each time that the test is performed and so the technique would automatically adjust accordingly. By applying methods of machine learning, it is possible for predictions of reliability to be made and techniques being customised automatically.

The final Chapters have explored a single locomotion task, using one interaction technique and future work needs to apply these methods to other interaction techniques and tasks. This work is currently in progress with a diagnostic test being developed based on gestures. Although, this is at a very early stage, I have started to develop a diagnostic test based on a ‘Simon Says’ game, see Figure 9.1. One of the lessons learned from the diagnostic test in this thesis is that it is not as motivating, engaging and exciting as playing a computer game might be. Therefore, using bright colours, pictures



Figure 9.1: A user with physical disabilities testing the ‘Simon Says’ based diagnostic test. The image sequences are showing the gesture pattern that he is begin asked to perform.

of favourite things and sounds, a gesture sequence is presented to the participant which, they are to repeat back using gaze. My expectation is that the results (based on the reliability metric), will allow a gesture based interaction technique to be customised for a user.

Since these experiments have taken place, parents and teachers have all seen increases in confidence and self-esteem of the four CP participants. The engagement that they have experienced and motivation in wanting to use the eye trackers, has helped to secure local authority funding for Tobii CEye eye tracking communication devices for two of the four participants. These two individuals are now using these on a daily basis within some of their classes. Additionally, one participant now has the confidence to learn how to use her wheelchair by herself using a single head switch. The school staff and her parents believe that this has been due to her feeling empowered by taking control of her own in-game character. Further to confidence building, one participant was so engaged with manouvering her character that the amount of positive verbal expression she made surprised even her mother. There are also therapeutic possibilities as evidenced by the participant who completely relaxed his tense body when his character walked into the water. These examples open up a new set of questions to explore how the game environment and state of the character can influence a person’s own mental and physical state.

Online games and environments are becoming more accepted as places of social interaction and sources for aiding education. Some of the most vulnerable young people in society have much to gain in accessing these worlds. With the cost of eye tracking reducing and open source alternatives, the role of dynamic accessibility and adaptive interaction techniques is essential in ensuring this group of people can access them.

Part IV

Appendices

Appendix A

Experiment 2, Supporting Data

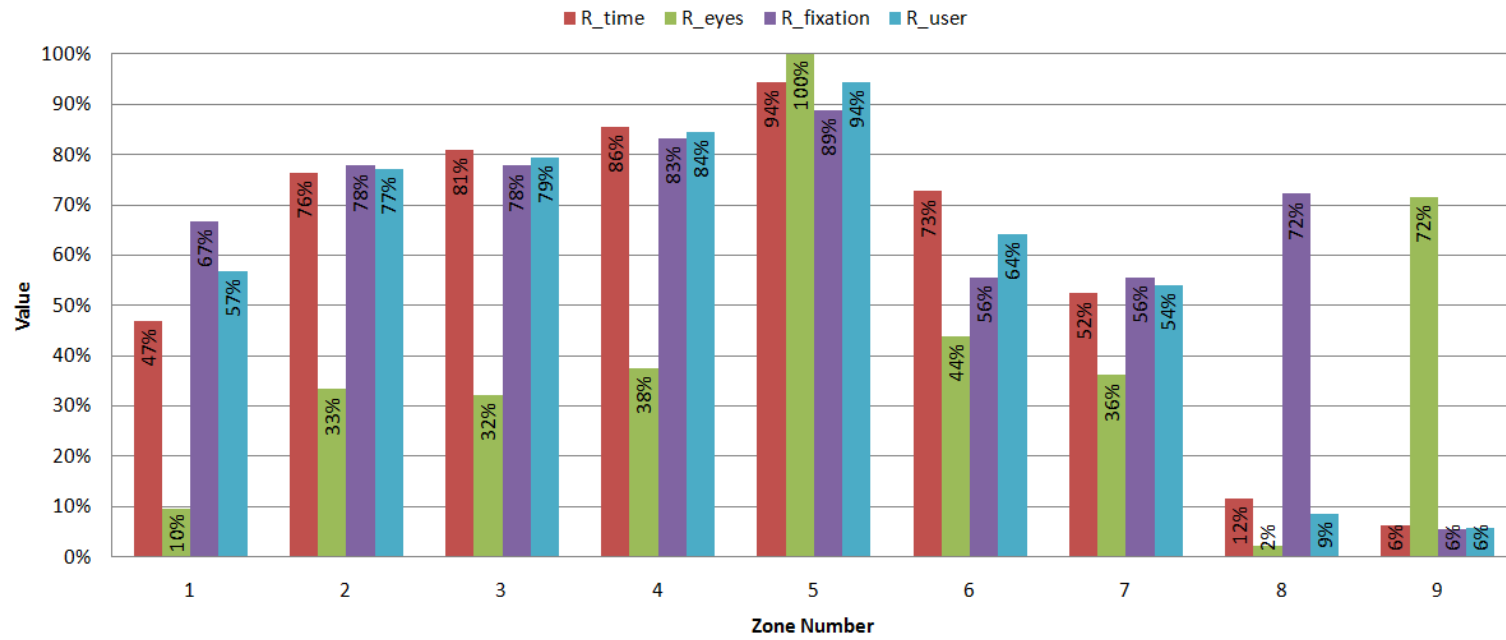


Figure A.1: Results from Experiment 2: GP-A1

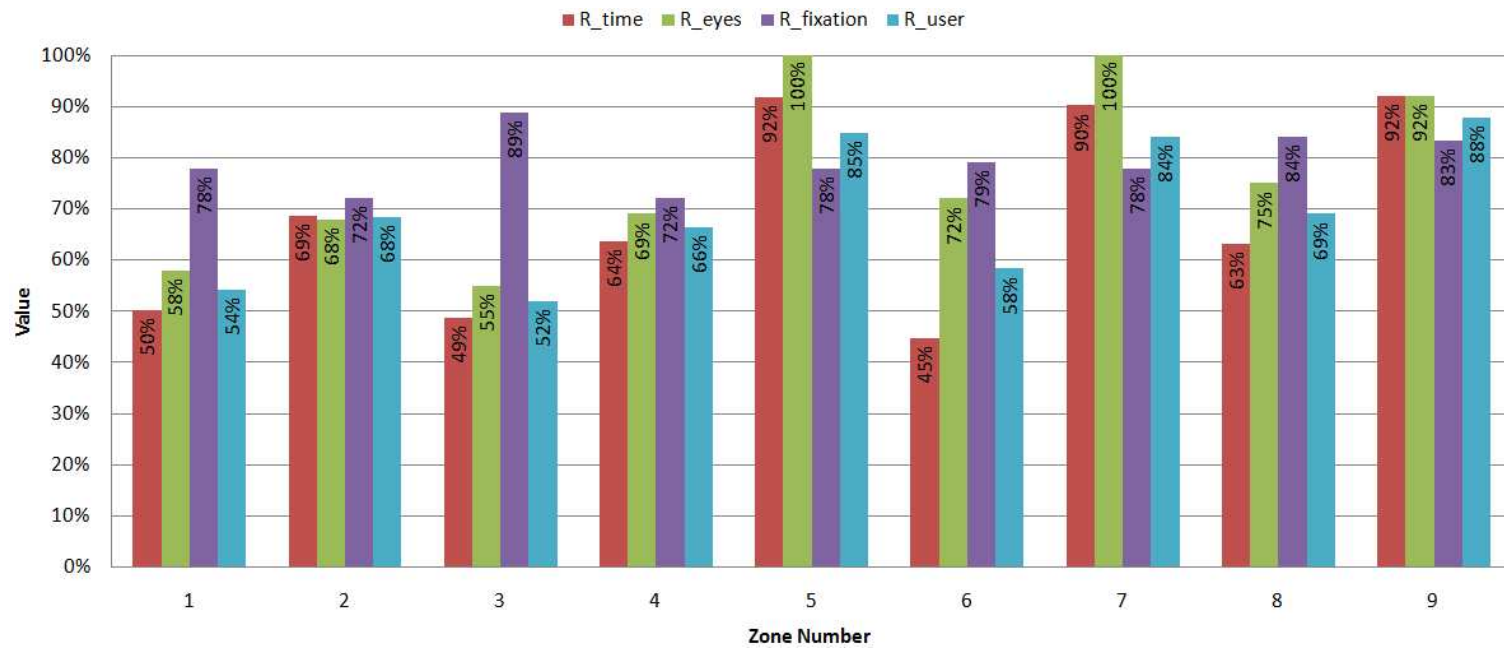


Figure A.2: Results from Experiment 2: GP-A2

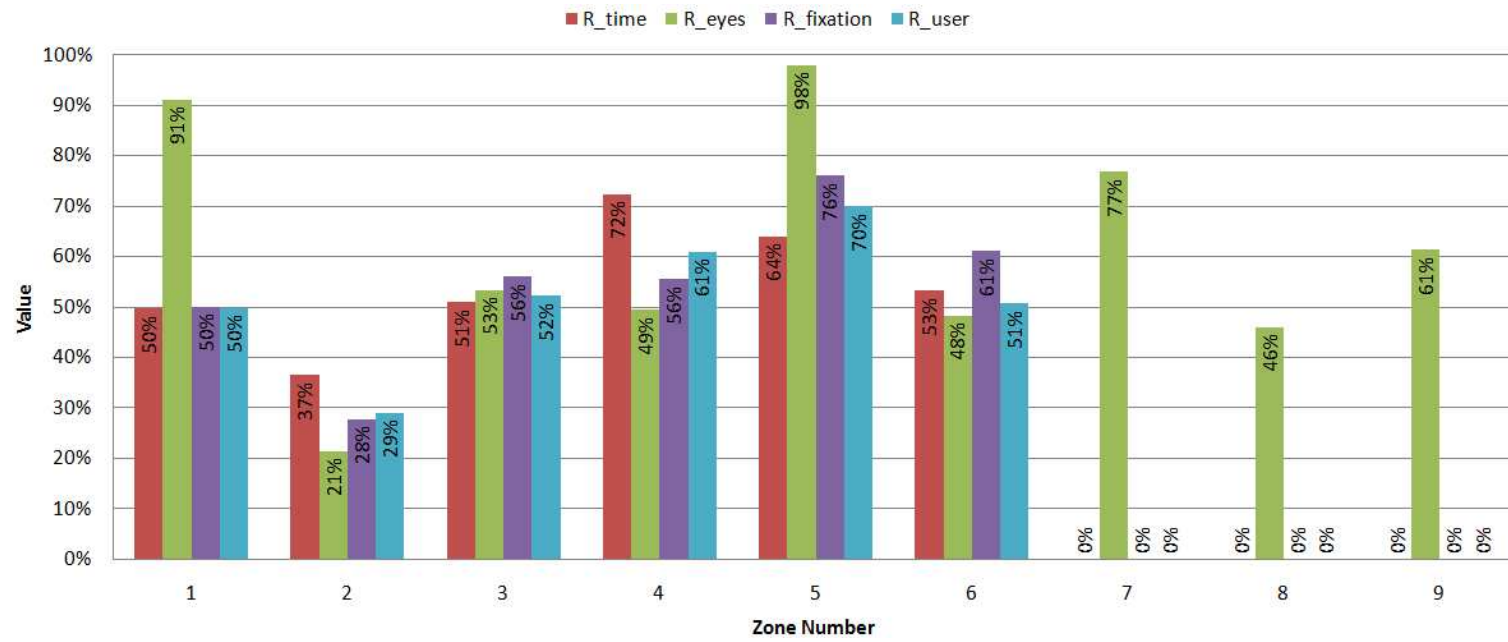


Figure A.3: Results from Experiment 2: GP-A3

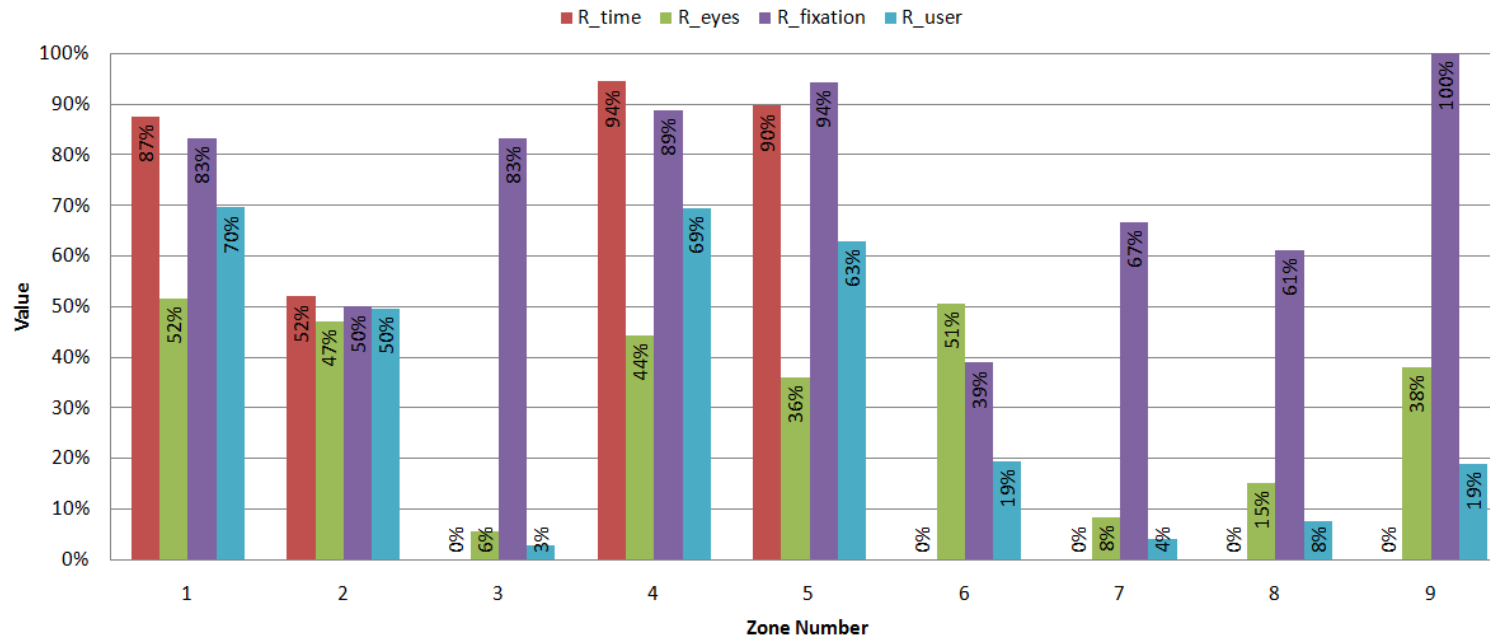


Figure A.4: Results from Experiment 2: GP-A4

Appendix B

Experiment 4, Supporting Data

B.1 Notes on Resampling

Resampling has been used as a method of statistical analysis for more than 50 years (Tukey, 1958). An explanation of the method used in Experiment 4: Group A v Group B is as follows (Figure B.1). The difference in median between Group A and Group B is 0.35 (35%). All of the samples are combined and shuffled into a new set of two groups of the original size and the difference in median between the two groups is calculated. This is repeated 10,000 times. See Howell (2001) for further information on resampling without replacement.

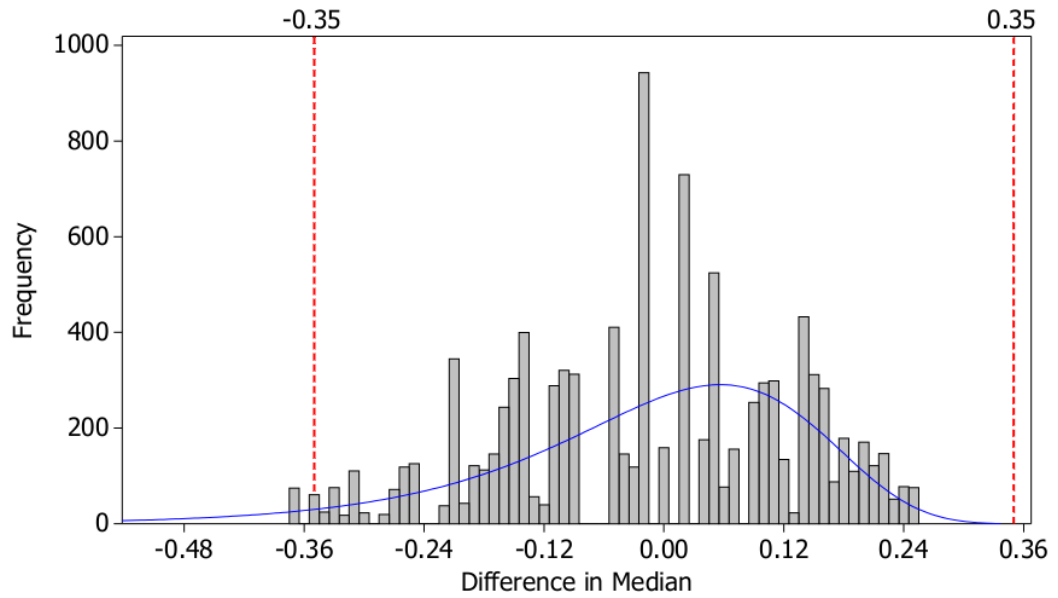


Figure B.1: Experiment 4: Data Resampled 10,000 times from Groups A and B

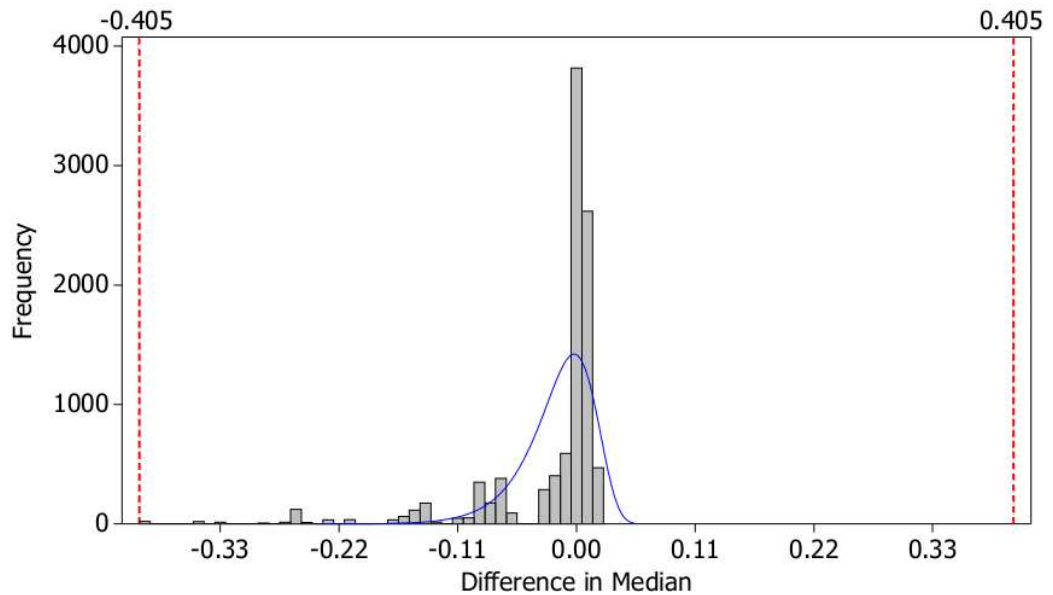


Figure B.2: Experiment 4: Data Resampled 10,000 times from Groups A and C

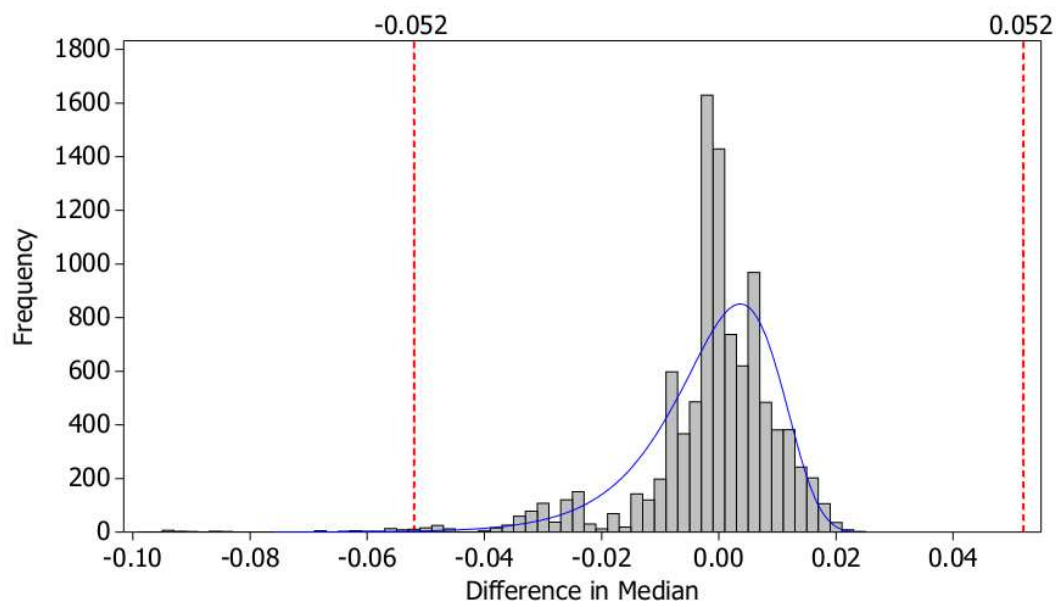


Figure B.3: Experiment 4: Data Resampled 10,000 times from Groups B and C

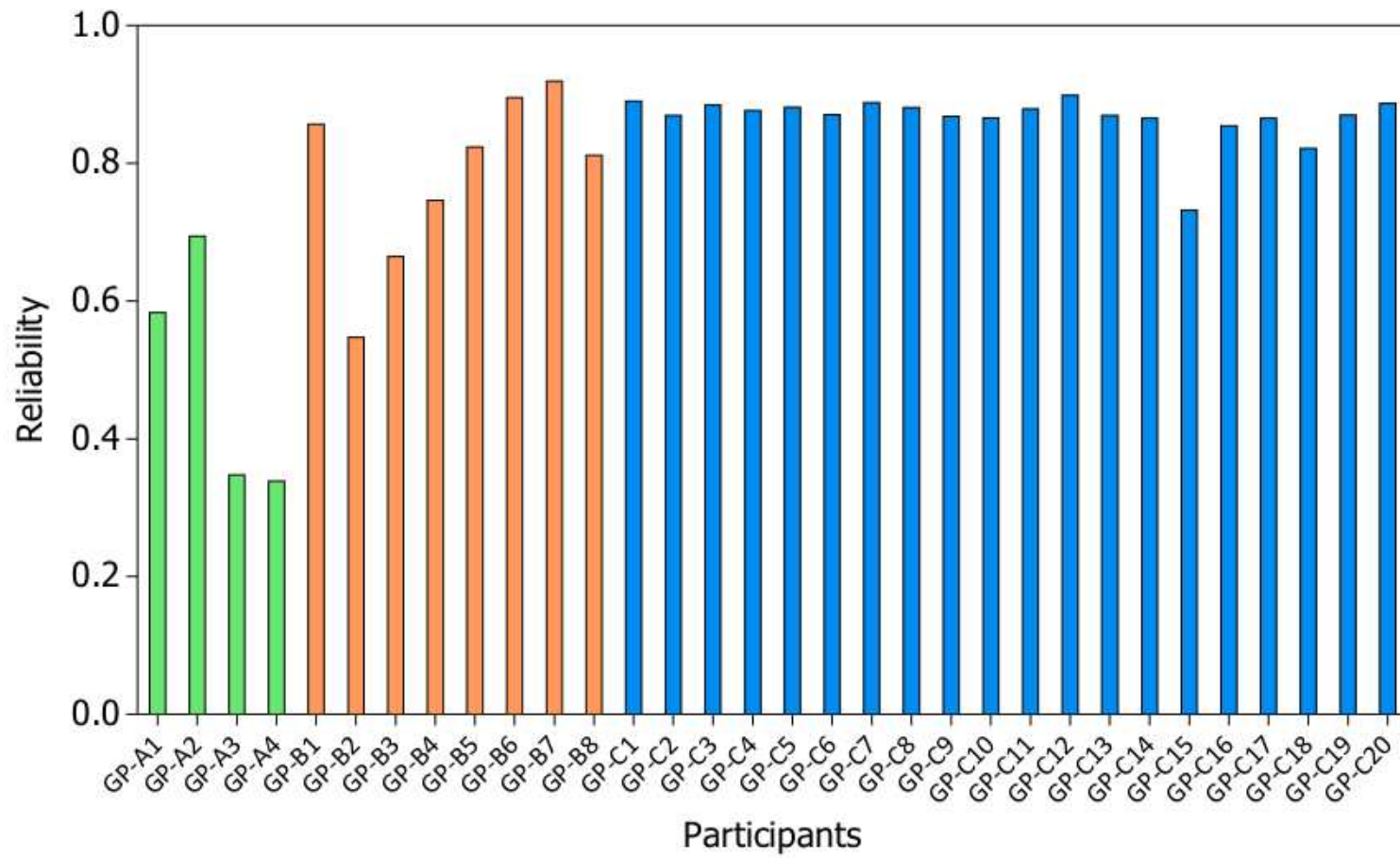


Figure B.4: Reliability values from the diagnostic tests in Experiment 2 and Experiment 4 for all participant Groups.

Appendix C

Experiment 5, Supporting Data

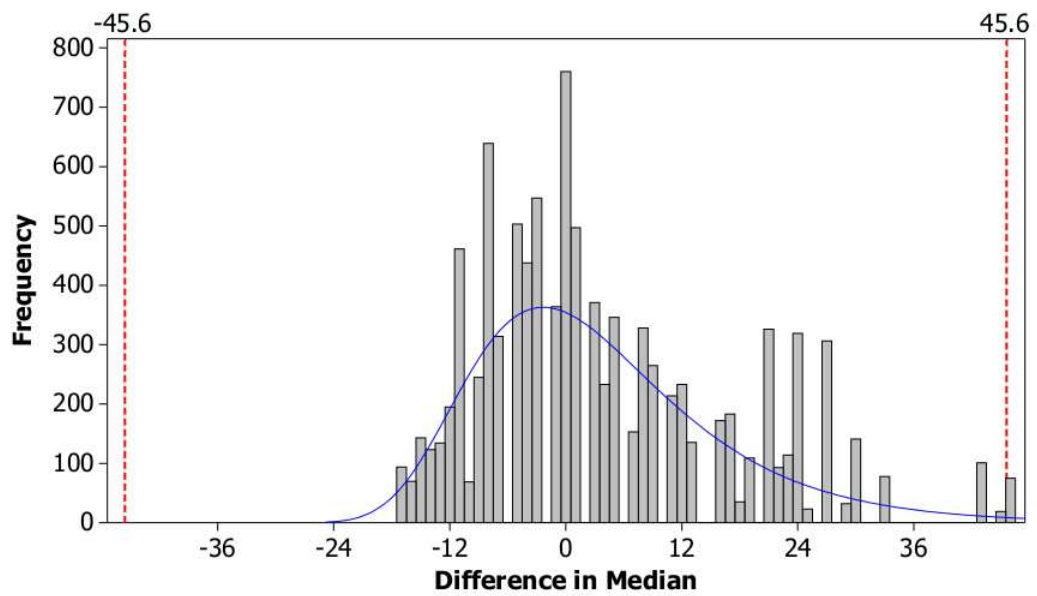


Figure C.1: Experiment 5: Data Resampled 10,000 times from Groups A and B

	Diagnostic Test		Time to Reach Way points (s)			
<i>Participant</i>	<i>success</i>	\overline{R}_{user}	<i>1st</i>	<i>2nd</i>	<i>3rd</i>	<i>Total</i>
GP-A1	94%	58%	86s	65s	DNF	DNF
GP-A2	94%	69%	43s	52s	33s	128s
GP-A3	55%	35%	75s	DNF	DNF	DNF
GP-A4	39%	34%	99s	DNF	DNF	DNF
GP-B1	100%	86%	43s	32s	21s	96s
GP-B2	78%	55%	50s	DNF	DNF	DNF
GP-B3	94%	66%	39s	51s	48s	138s
GP-B4	91%	75%	30s	25s	27s	82s
GP-B5	100%	82%	28s	23s	25s	77s
GP-B6	100%	90%	33s	27s	34s	94s
GP-B7	100%	92%	26s	28s	20s	74s
GP-B8	100%	81%	38s	23s	22s	83s
GP-C1	100%	89%	26s	23s	20s	69s
GP-C2	100%	87%	32s	36s	21s	89s
GP-C3	100%	88%	33s	39s	19s	92s
GP-C4	100%	88%	27s	20s	20s	67s
GP-C5	100%	88%	29s	24s	14s	67s
GP-C6	100%	87%	28s	22s	19s	69s
GP-C7	100%	89%	29s	43s	50s	122s
GP-C8	100%	88%	27s	20s	22s	69s
GP-C9	100%	87%	29s	33s	22s	84s
GP-C10	100%	87%	27s	24s	54s	105s
GP-C11	100%	88%	32s	23s	25s	80s
GP-C12	100%	90%	30s	23s	25s	78s
GP-C13	100%	87%	40s	30s	24s	94s
GP-C14	94%	87%	DNF	DNF	DNF	DNF
GP-C15	100%	73%	36s	30s	25s	91s
GP-C16	100%	85%	31s	23s	23s	78s
GP-C17	100%	87%	39s	21s	21s	81s
GP-C18	100%	82%	49s	DNF	DNF	DNF
GP-C19	100%	87%	33s	27s	46s	96s
GP-C10	100%	89%	42s	DNF	DNF	DNF

Table C.1: Experiment 3 and 5: Summary of task completion times and associated results from the diagnostic test (DNF = Did Not Finish)

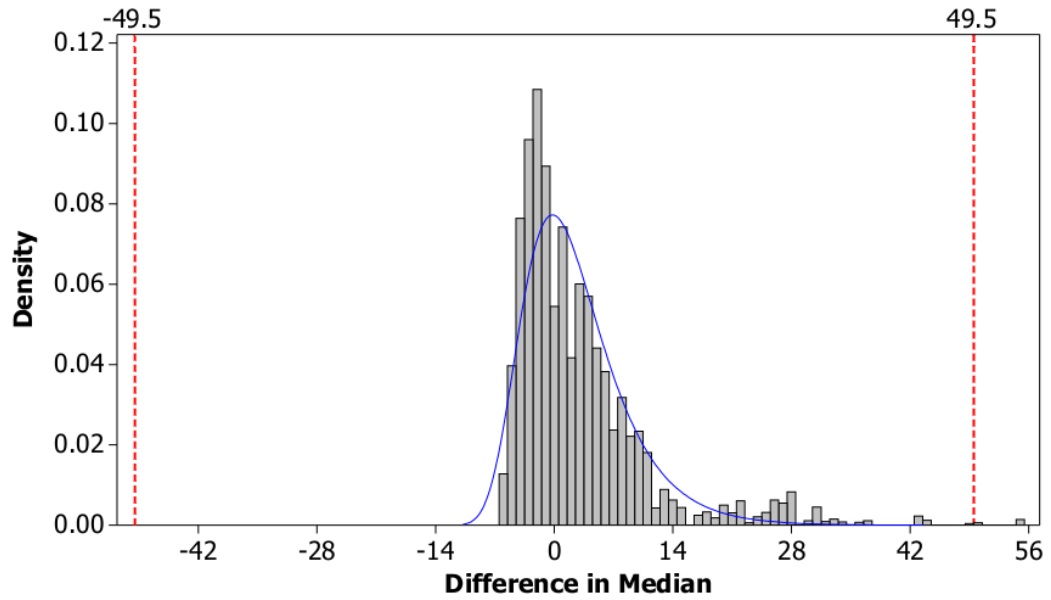


Figure C.2: Experiment 5: Data Resampled 10,000 times from Groups A and C

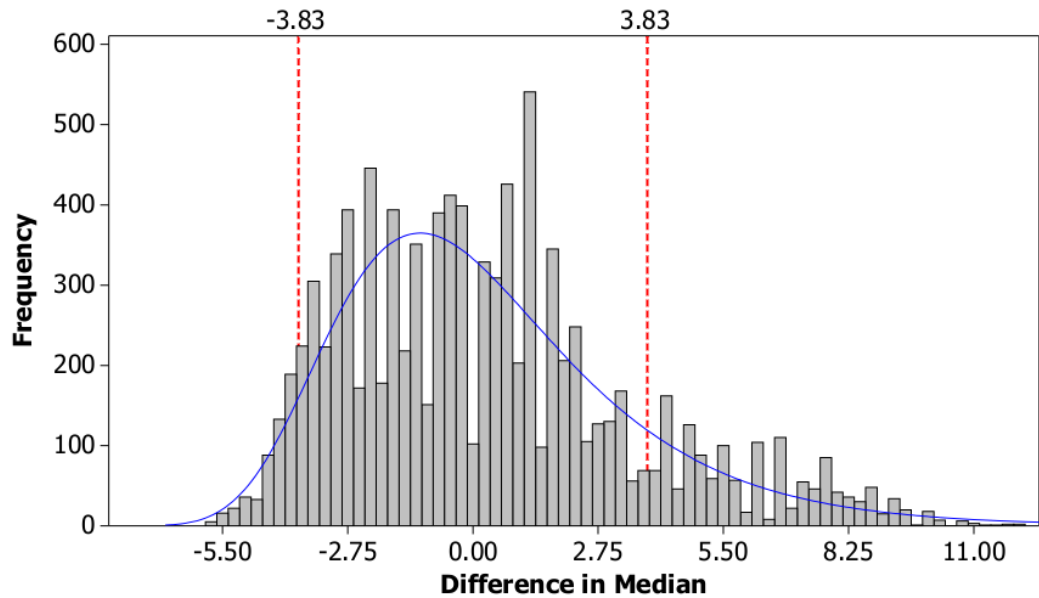


Figure C.3: Experiment 5: Data Resampled 10,000 times from Groups B and C

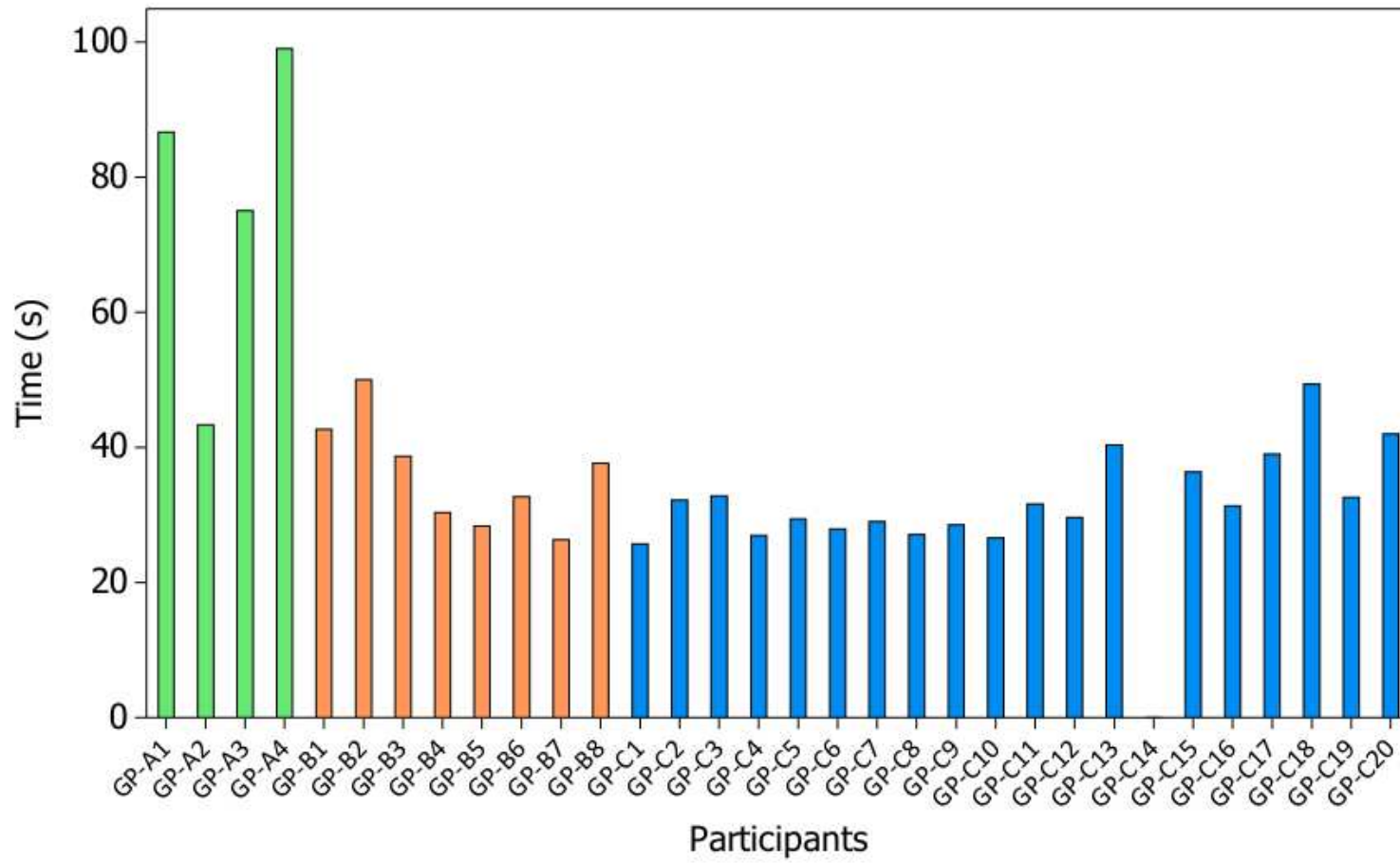


Figure C.4: Time to reach the first waypoint from Experiment 3 and Experiment 5 for all participant Groups.

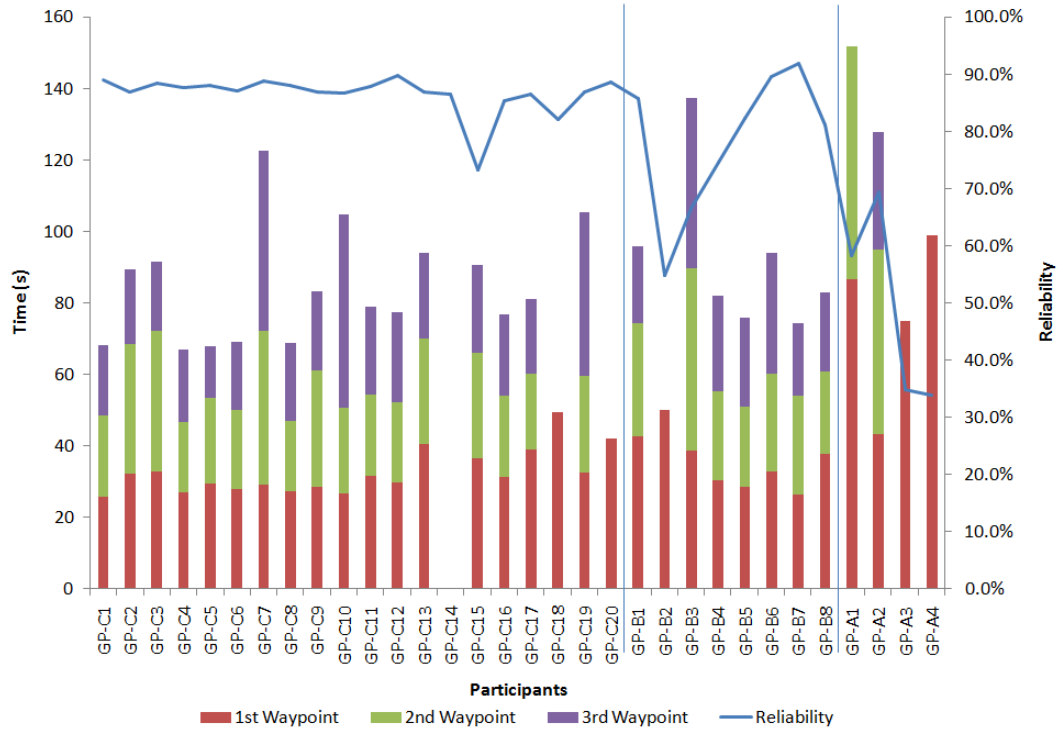


Figure C.5: Task completion times for all Groups with reliability overlaid

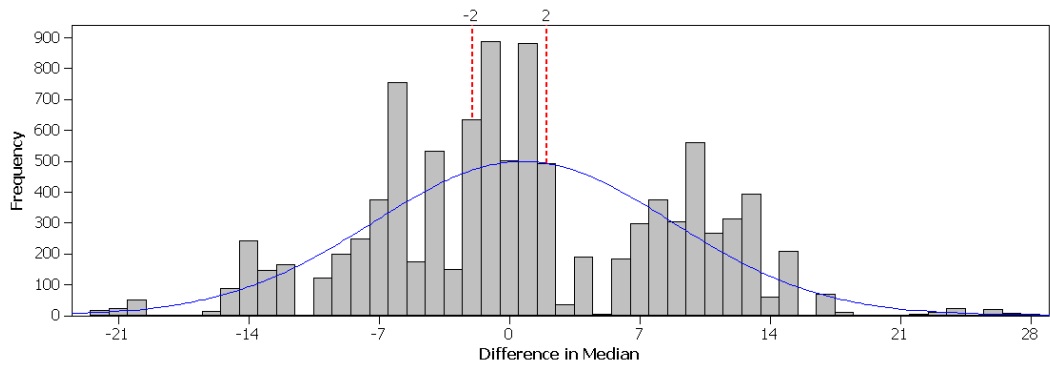


Figure C.6: Experiment 5: Data Resampled 10,000 times from Groups B and C. This is based on total time.

References

- Annett, J. (2003), ‘Hierarchical task analysis’, *Handbook of cognitive task design* **2**, 17–35. 114
- Barreto, A. B., Scargle, S. D. and & Adjouadi, M. (2000), ‘A practical emg-based human-computer interface for users with motor disabilities’, *Journal of Rehabilitation Research and Development* **37**(1), 53–64. 28
- Bates, R. (2002), Computer input device selection methodology for users with high-level spinal cord injuries, *in* ‘Proceedings of the 1st Cambridge Workshop on Universal Access and Assistive Technology (CWUAAT)’. 4
- Bates, R., Istance, H. and Vickers, S. (2008), Gaze interaction with virtual on-line communities: Levelling the playing field for disabled users, *in* ‘Designing Inclusive Futures’, Springer-Verlag, pp. 149–162. 4
- Bee, N. and Andre, E. (2008), Writing with your eye: A dwell time free writing system adapted to the nature of human eye gaze, *in* ‘Perception in Multimodal Dialogue Systems, LNCS’, pp. 111–122. 30, 35, 45
- Berk, L. E. (2009), *Child Development*, 8th edn, Pearson. 25
- Bigham, J. P., Kaminsky, R. S., Ladner, R. E., Danielsson, O. M. and Hempton, G. L. (2006), Webinsight: Making web images accessible, *in* ‘8th international ACM SIGACCESS conference on Computers and accessibility; ASSETS ’06’. 45
- Blizzard (2010), ‘World of warcraft subscriber base reaches 12 million worldwide’, Internet (Retrieved March 2011).
URL: <http://us.blizzard.com/en-us/company/press/pressreleases.html?101007> 3, 51

REFERENCES

- Brotherson, S. (2009), ‘Supporting physical growth and development in young children’, *Bright Beginnings* **9**. 25
- Bulling, A., Roggen, D. and Tröster, G. (2009), Wearable eog goggles: eye-based interaction in everyday environments, *in* ‘CHI ’09: 27th international conference extended abstracts on Human factors in computing systems’. 17, 18
- Bushby, K. M. D. and Anderson, L. V. B. (2001), *Muscular dystrophy: methods and protocols*, Humana Press. 19
- Buxton, W. (1990), A three-state model of graphical input, *in* ‘Human-Computer Interaction - INTERACT ’90’, pp. 449–456. 110
- Card, S., Moran, T. and Newell, A. (1983), *The psychology of human-computer interaction*, Hillsdale, N.J. : L. Erlbaum Associates. 105, 110
- Carpenter, R. H. S. (1988), *Movements of the Eyes*, London: Pion. 14
- Carter, S., Hurst, A., Mankoff, J. and Li, J. (2006), Dynamically adapting guis to diverse input devices, *in* ‘Eighth international conference on Assistive technologies; ASSETS’. 45
- Castellina, E. and Corno, F. (2007), Accessible web surfing through gaze interaction, *in* ‘Proceedings of the 3rd Conference on Communication by Gaze Interaction: COGAIN’. 30, 41
- Davson, H. (1990), *Physiology of the Eye*, 5th ed edn, Pergamon. 14
- Dawe, M. (2006), Desperately seeking simplicity: how young adults with cognitive disabilities and their families adopt assistive technologies, *in* ‘SIGCHI conference on Human Factors in computing systems’. 5, 45
- Dix, A., Finlay, J., Abowd, G. and Beale, R. (1998), *Human-Computer Interaction*, 2nd ed edn, London : Prentice Hall. 104
- Donegan, M., Oosthuizen, L., Bates, R., Daunys, G., Hansen, J., Joos, M., Majaranta, P. and Signorile, I. (2005), ‘D3.1 user requirements report with observations of difficulties users are experiencing’, *Communication by Gaze Interaction (COGAIN)* . 5, 22, 195

- Donegan, M., Oosthuizen, L., Bates, R., Istance, H., Holmqvist, E. and Lundälv, M. (2006), ‘D3.3 report of user trials and usability studies’, *Communication by Gaze Interaction (COGAIN)* . 28
- Dorr, M., Bohme, M., Martinetz, T. and Barth, E. (2007), Gaze beats mouse: A case study, in ‘Proceedings of the 3rd Conference on Communication by Gaze Interaction: COGAIN’. 30, 43
- Drewes, H. and Schmidt, A. (2007), Interacting with the computer using gaze gestures, in ‘Proceedings of INTERACT 2007’. 30, 31, 32, 41, 44, 45
- Duchowski, A. (2003), *Eye Tracking Methodology: Theory and Practice*, Springer. 12, 13, 14
- Eye diagram* (2010), Internet (Retrieved July 2010).
URL: <http://www.nei.nih.gov/photo/> 11
- Fajen, B. R. and Warren, W. H. (2003), ‘Behavioral dynamics of steering, obstacle avoidance, and route selection’, *Journal of Experimental Psychology Human Perception and Performance* **29**(2), 343–362. 97
- Fono, D. and Vertegaal, R. (2005), Eyewindows: Evaluation of eye-controlled zooming windows for focus selection, in W. Kellogg, ed., ‘Conference on human factors in computing systems; CHI 2005: technology, safety, community’, New York, Portland, OR., pp. 151–160. 28
- Friedman, M. B., Kiliansky, G., Dzamura, M. and D, A. (1982), ‘The eyetracker communication system’, *Johns Hopkins APL Technical Digest* **3**(3), 250–252. 4
- Gajos, K. Z. (2008), Automatically Generating Personalized User Interfaces, PhD thesis, University of Washington, Seattle. 45, 46
- Gajos, K. Z., Wobbrock, J. O. and Weld, D. S. (2008), Improving the performance of motor-impaired users with automatically-generated, ability-based interfaces, in ‘26th annual SIGCHI conference on Human factors in computing systems (CHI ’08)’. 45
- Gilb, T. (1984), The “impact analysis table” applied to human factors design, in ‘Interact ’84, First IFIP Conference on Human-Computer Interaction’, pp. 97–101. 80

- Gips, J. and Olivieri, P. (1996), Eagleeyes: An eye control system for persons with disabilities, *in* ‘The 11th International Conference on Technology and Persons with Disabilities’, Los Angeles, California. 17, 18, 24, 30, 39, 40
- Guimbretiere, F., Dixon, M. and Hinckley, K. (2007), ‘Experiscope: an analysis tool for interaction data’, pp. 1333–1342. 104, 105, 109
- Hand, C. (1997), ‘A survey of 3d interaction techniques’, *Computer Graphics Forum* **16**(5), 269–282. 54, 61
- Hansen, D. W., Skovsgaard, H. H., Hansen, J. P. and Møllenbach, E. (2008), Noise tolerant selection by gaze-controlled pan and zoom in 3d, *in* ‘Symposium on Eye Tracking Research & Applications (ETRA)’. 30, 37
- Hansen, J. P., A.S., J., Hansen, D., Itoh, K. and Mashino, S. (2003), Command without a click: Dwell time typing by mouse and gaze selections, *in* ‘9th IFIP International Conference on Human Computer Interaction (INTERACT)’. 29, 30
- Hansen, J. P., Hansen, D., Johansen, A. and Elvesjö, J. (2005), Mainstreaming gaze interaction towards a mass market for the benefit of all, *in* ‘3rd international conference on universal access in human-computer interaction (HCII 2005)’. 5
- Hansen, J. P., Lund, H., Aoki, H. and Itoh, K. (2006), Gaze communication systems for people with als, *in* ‘17th International Symposium on ALS/MND’. 4, 130
- Hartson, H. R., Siochi, A. C. and Hix, D. (1990), ‘The uan: a user-oriented representation for direct manipulation interface designs’, *ACM Trans. Inf. Syst.* **8**, 181–203. 108
- Heikkila, H. and Raiha, K. J. (2009), ‘Speed and accuracy of gaze gestures’, *Journal of Eye Movmen Research* **3**(2), 1–14. 30, 40, 41, 44, 45
- Hori, J., Sakano, K., Miyakawa, M. and Saitoh, Y. (2006), ‘Eye movement communication control system based on eog and voluntary eye blink’, *LNCS* **4061**, 950–953. 17
- Hornof, A. J. and Cavender, A. (2005), Eyedraw: Enabling children with severe motor impairments to draw with their eyes, *in* ‘Conference on human factors in computing systems; CHI 2005: technology, safety, community’. 25, 30, 39, 40

- Howell, D. C. (2001), *Statistical Methods for Psychology*, 5th edn, Wadsworth Publishing Co Inc. 206
- Huckauf, A., Goettel, T., Heinbockel, M. and Urbina, M. (2005), What you don't look at is what you get: anti-saccades can reduce the midas touch-problem, in 'APGV '05: Proceedings of the 2nd symposium on Applied perception in graphics and visualization', pp. 170–170. 30
- ISO (2000), *ISO 9241-9:2000. Ergonomic requirements for office work with visual display terminals (VDTs) – Part 9: Requirements for non-keyboard input devices*, International Organization for Standardization. 46
- Isokoski, P. (2000), Text input methods for eye trackers using off-screen targets, in 'ETRA '00: Proceedings of the 2000 symposium on Eye tracking research & applications', pp. 15–21. 35
- Isokoski, P., Joos, M., Martin, M. and Spakov, O. (2009), 'Gaze controlled games', *Universal Access in the Information Society* . 30, 42, 43
- Isokoski, P. and Martin, B. (2006), Eye tracker input in first person shooter games:, in 'Communication by Gaze Interaction; COGAIN 2006'. 5
- Istance, H., Bates, R., Hyrskykari, A. and Vickers, S. (2008), Snap clutch, a moded approach to solving the midas touch problem, in 'Proceedings of the 2008 Symposium on Eye Tracking Research and Applications; ETRA 2008'. 66, 70, 72
- Istance, H., Hyrskykari, A., Immonen, L., Mansikkamaa, S. and Vickers, S. (2010), Designing gaze gestures for gaming: an investigation of performance, in 'Proceedings of the 2010 Symposium on Eye Tracking Research and Applications; ETRA 2010'. 30, 31
- Istance, H., Hyrskykari, A., Vickers, S. and Chaves, T. (2009), For your eyes only: Controlling 3d online games by eye-gaze, in 'Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: Part I'. 73
- Istance, H. O., Spinner, C. and Howarth, P. A. (1996), Providing motor-impaired users with access to standard graphical user interface software via eye-based interaction,

REFERENCES

- in* ‘International conference on disability, virtual reality and associated technologies’, p. 13. 30, 31, 34, 35
- Jacob, R. (1993), Eye-movement-based human-computer interaction techniques: Toward non-command interfaces, *in* ‘Advances in Human-Computer Interaction’, Vol. 4, pp. 151–190. 28, 41
- Jacob, R. J. K. (1990), What you look at is what you get: eye movement-based interaction techniques, *in* ‘CHI ’90: Proceedings of the SIGCHI conference on Human factors in computing systems’, pp. 11–18. 29
- Javal, L. E. (1907), *Physiologie de la lecture et de l'écriture*, Bibliography in *Annales oculistique*. 10
- Johansen, A. S., Hansen, J. P., Hansen, D. W., Itoh, K. and Mashino, S. (2003), Language technology in a predictive, restricted on-screen keyboard with dynamic layout for severely disabled people, *in* ‘TextEntry ’03: Proceedings of the 2003 EACL Workshop on Language Modeling for Text Entry Methods’, pp. 59–66. 4, 30, 34
- Jönsson, E. (2005), If looks could kill - an evaluation of eye tracking in computer games, Master’s thesis, Royal Institute of Technology, Stockholm. 30, 43
- Just, M. A. and Carpenter, P. A. (1980), ‘A theory of reading: From eye fixations to comprehension’, *Psychological Review* **87**(4), 329–354. 15
- Kate, J. t., Frietman, E., Willems, E., Ter Haar Romeny, B. and Tenkink, E. (1979), Eye-switch controlled communication aids, *in* ‘Proceedings of the 12th International Conference on Medical & Biological Engineering’. 4
- Kiyohiko, A., O, K., Shoichi, O. and Minoru, O. (2007), ‘A web browsing system by eye-gaze input’, *IEEEJ Transactions on Electronics, Information and Systems* **7**, 1036–1042. 30, 41
- Krauzlis, R. J. and Lisberger, S. G. (1994), ‘Temporal properties of visual motion signals for the initiation of smooth pursuit eye movements in monkeys.’, *J Neurophysiol* **72**(1), 150–162. 14

REFERENCES

- Kumar, M., Paepcke, A. and Winograd, T. (2007), Eyepoint: Practical pointing and selection using gaze and keyboard, *in* ‘Proceedings of SIGCHI Conference on Human Factors in Computing Systems: CHI ’07’. 28
- Kumar, M. and Winograd, T. (2007), Gaze-enhanced scrolling techniques, *in* ‘UIST ’07: Proceedings of the 20th annual ACM symposium on User interface software and technology’, ACM, New York, NY, USA, pp. 213–216. 30, 41, 42
- Lankford, C. (2000), Effective eye-gaze input into windows, *in* ‘Proceedings of the 2000 Symposium on Eye Tracking Research and Applications; ETRA 2000’. 4, 30, 31
- Li, D. and Parkhurst, D. J. (2006), Open-source software for real-time visible-spectrum eye tracking, *in* ‘Proceedings of the 2nd Conference on Communication by Gaze Interaction: COGAIN’. 19
- Linden-Labs (2011), ‘The second life economy’, Internet (Retrieved March 2011).
URL: <http://community.secondlife.com/t5/Featured-News/The-Second-Life-Economy-in-Q4-2010/> 2, 51
- Mackinlay, J. D., Card, S. K. and Robertson, G. G. (1990), Rapid controlled movement through a virtual 3d workspace, *in* ‘SIGGRAPH ’90: 17th annual conference on Computer graphics and interactive techniques’, pp. 171–176. 98
- Majaranta, P. (2009), Text Entry by Eye Gaze, PhD thesis, University of Tampere. 30
- Majaranta, P., MacKenzie, I. S., Aula, A. and Raiha, K. J. (2006), ‘Effects of feedback and dwell time on eye typing speed and accuracy’, *Universal Access in the Information Society* **5**(2), 199–208. 29
- Majaranta, P. and Raiha, K. J. (2002), Twenty years of eye typing: Systems and design issues, *in* ‘Proceedings of the 2002 Symposium on Eye Tracking Research and Applications; ETRA 2002’. 5, 29, 33, 34, 35
- Mankoff, J., Moore, M., Batra, U. and Dey, A. (2002), Web accessibility for low bandwidth input, *in* ‘Proceedings of the fifth international ACM conference on Assistive technologies; ASSETS’. 45

REFERENCES

- MDA (2010), ‘Duchenne muscular dystrophy’, Internet (Retrieved March 2011).
URL: <http://www.mda.org/disease/dmd.html> 20
- Miniotas, D. (2000), Application of fitts’ law to eye gaze interaction, *in* ‘CHI ’00 extended abstracts on Human factors in computing systems’. 154
- Møllenbach, E., Lillholm, M., Gail, A. and Hansen, J. P. (2010), Single gaze gestures, *in* ‘ETRA ’10: Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications’, pp. 177–180. 30, 44, 45
- Nantais, T., Shein, F. and Treviranus, J. (1994), ‘A predictive selection technique for single digit typing with a visual keyboard’, *IEEE Transactions on Rehabilitation Engineering* **2**(3). 34
- Nielsen, J. (1993), ‘Noncommand user interfaces’, *Communications of the ACM* **36**(4), 83–99. 28
- Nievergelt, J. and Weydert, J. (1987), ‘Human-computer interaction: a multidisciplinary approach’, pp. 438–441. 65
- Perlin, K. (1998), Quikwriting: continuous stylus-based text entry, *in* ‘11th Annual ACM Symposium on User interface Software and Technology (UIST ’98)’. 35
- RCO and RNIB (2010), ‘Understanding nystagmus’, Internet (Retrieved March 2011). Royal College of Ophthalmologists and Royal National Institute of Blind People.
URL: <http://www.rnib.org.uk/eyehealth/eyeconditions/eyeconditionsdn/Pages/nystagmus.aspx> 24
- Renzi, E. D., Colombo, A., Faglioni, P. and Gibertoni, M. (1982), ‘Conjugate gaze paresis in stroke patients with unilateral damage’, *Archives of Neurology* **39**(8), 482–486. 24
- Robinson, D. (1968), The oculomotor control system: A review, *in* ‘Proceedings of the IEEE’, pp. 1032–1049. 16
- Rosenbaum, P., Paneth, N., Leviton, A., Goldstein, M., Bax, M., Damiano, D., Dan, B. and Jacobsson, B. (2007), ‘A report: the definition and classification of cerebral palsy’, *Developmental medicine and child neurology* **49**, 8–14. 20

- Ruddle, R. A. and Lessels, S. (2006), ‘Three levels of metric for evaluating wayfinding’, *Presence* **15**, 637–654. 98
- Salvucci, D. D. (1999), Inferring intent in eye-based interfaces: tracing eye movements with process models, *in* ‘CHI ’99: SIGCHI conference on Human factors in computing systems’, pp. 254–261. 34
- San Agustin, J., Hansen, J. P. and Mateo, J. (2008), Hands-free selection by combining gaze pointing and emg selection, *in* ‘Scandinavian Workshop on Applied Eye Tracking’. 28
- San Agustin, J., Skovsgaard, H., Hansen, J. P. and Hansen, D. W. (2009), Low-cost gaze interaction: ready to deliver the promises, *in* ‘27th international Conference Extended Abstracts on Human Factors in Computing Systems’. 5, 19
- Satty, T. L. (1990), *Multicriteria decision making : the analytic hierarchy process : planning, priority setting resource allocation*, RWS Publications. 118
- Skovsgaard, H., Mateo, J. C. and Flach, M. F. H. J. P. (2010), Small-target selection with gaze alone, *in* ‘2010 Symposium on Eye Tracking Research and Applications; ETRA 2010’. 4, 31
- Skovsgaard, H., P, H. J. and Mateo, J. C. (2008), How can tiny buttons be hit using gaze only?, *in* ‘4th Conference on Communication by Gaze Interaction (COGAIN)’. 30, 31
- Smith, J. D. and Graham, T. C. N. (2006), Use of eye movements for video game control, *in* ‘Proceedings of the 2006 International Conference on Advances in Computer Entertainment Technology: ACE ’06’. 5, 30, 43
- Snowden, R., Thompson, P. and Troscianko, T. (2006), *Basic Vision: An introduction to visual perception*, Oxford University Press. 15
- Stampe, D. M. and Reingold, E. M. (1995), ‘Selection by looking: A novel computer interface and its application to psychological research’, *Eye movement research: Mechanisms, processes and applications* pp. 467–478. 28, 29

REFERENCES

- Starker, I. and Bolt, R. A. (1990), A gaze-responsive self-disclosing display, *in* 'CHI '90: Proceedings of the SIGCHI conference on Human factors in computing systems', pp. 3–10. 30, 44
- Sternberg, R. (2003), *Cognitive Psychology*, 3rd edn, Thomson Wadsworth. 154
- Tchalenko, J. (2001), 'Free-eye drawing', *Point: Art and Design Research Journal* **11**, 36–41. 30, 39, 41
- Templeman, J. N., Denbrook, P. S. and Sibert, L. E. (1999), 'Virtual locomotion: Walking in place through virtual environments', *Presence* **8**(6), 598–617. 98
- Tukey, J. W. (1958), 'Bias and confidence in not quite large samples', *The Annals of Mathematical Statistics* **29**(614). 206
- Vertegaal, R. (2008), A fitts law comparison of eye tracking and manual input in the selection of visual targets, *in* 'ICMI '08 10th international conference on Multimodal interfaces'. 154
- Vickers, S., Istance, H. and Smalley, M. (2010), Eyeguitar: Making rhythm based music video games accessible using only eye movements, *in* 'International Conference on Advances in Computer Entertainment Technology: ACE '10'. 124
- Ward, D. J., Blackwell, A. F. and MacKay, D. J. C. (2000), Dasher - a data entry interface using continuous gestures and language models, *in* 'UIST 2000: The 13th Annual ACM Symposium on User Interface Software and Technology'. 30, 37
- Ward, D. J. and Mikaelian, H. H. (2002), 'Fast hands-free writing by gaze direction', *Nature* **418**, 838. 5, 37
- Ware, C. and Mikaelian, H. (1987), 'An evaluation of an eye tracker as a device for computer input'. 27, 28
- Wobbrock, J. O., Myers, B. A. and Kembel, J. A. (2003), Edgewise: A stylus-based text entry method designed for high accuracy and stability of motion, *in* '16th Annual Symposium on User Interface Software and Technology (UIST)'. 32, 36

REFERENCES

- Wobbrock, J. O., Rubinstein, J., Sawyer, M. W. and Duchowski, A. T. (2008), Longitudinal evaluation of discrete consecutive gaze gestures for text entry, *in* 'ETRA '08: 2008 symposium on Eye tracking research and applications', pp. 11–18. 30, 36, 45
- Young, L. R. and Sheena, D. (1975), 'Methods and designs - survey of eye movement recording methods', *Behaviour Research Methods and Instrumentation* **7**, 397–429. 15, 17
- Zhai, S., Morimoto, C. and Ihde, S. (1999), Manual and gaze input cascaded (magic) pointing, *in* 'CHI 99: the CHI is the limit, human factors in computing systems', pp. 246–253. 4, 30, 33
- Zhang, X. and MacKenzie, I. S. (2007), Evaluating eye tracking with iso 9241 - part 9, *in* 'Human-Computer Interaction, Part III, HCII 2007'. 154